

Cuantificación de la no invariancia y su aplicación en Estadística

Gloria García García

Departament d'Estadística
Universitat de Barcelona

Programa de doctorado
Probabilitats i Estadística
Bienio 1995–1997

Memoria presentada por Gloria M.
García García para optar al grado
de Doctor en Matemáticas por la
Universitat de Barcelona.

Certifico que esta tesis ha estado
realizada bajo mi dirección por
GLORIA M. GARCÍA GARCÍA en
el Departament d'Estadística de
la Universitat de Barcelona,

Dr. Josep María Oller i Sala
Catedrático de Universidad,
Departament d'Estadística,
Universitat de Barcelona.

A mi madre.

Puede ser muy arduo definir la belleza matemática, pero eso mismo sucede con cualquier otro tipo de belleza. Quizá no conocamos exactamente qué entendemos por un poema hermoso, pero no nos impide reconocerlo cuando lo leemos.”

G.H. Hardy
Apología de un matemático

Agradecimientos

Al Dr. Josep Maria Oller por haber aceptado ser mi director de tesis. También le quiero agradecer su dedicación, su tiempo y el proporcionarme todos los medios a su alcance para que esta tesis pudiera realizarse.

A la Dra. Marta Sanz por sus excelentes clases y por enseñarme cómo disfrutar del aprendizaje.

Al Dr. Josep Fortiana por sus comentarios e interés a lo largo de todos estos años y, en general, a los miembros del Departament d'Estadística de la Universitat de Barcelona.

A mi familia, claro. A mi madre, por su paciencia infinita y a mi hermano que siempre está ahí, apoyándome a su manera.

A mis amigos. A Anna por los 13 años de interesantes conversaciones matemáticas y no matemáticas. A Anna Pujol y por supuesto a Pol por recordarme, de vez en cuando, las cosas verdaderamente importantes. Y a Gisela y a Jose por su amistad con mayúsculas. Ya saben Aurea, Ferran, Jaume y Toni que para ellos también va mi agradecimiento.

Finalmente quiero dar las gracias a Manel por todo lo que nos hemos reído juntos y porque sin él esta tesis probablemente no se hubiera podido realizar.

Introducción

En Estadística el término *invariancia* puede ser entendido como sinónimo de simetría. Muchos problemas estadísticos son invariantes frente a la acción de un grupo de transformaciones, condición que hace natural restringir nuestra atención a reglas de decisión que también sean invariantes. Este principio de selección es ampliamente aceptado, tal y como puede consultarse en Lehmann [28], Giri [19] y [20], Muirhead [30] o bien en Eaton [17], y obviarlo puede llevar a soluciones incoherentes con la naturaleza del problema a resolver.

Si disponemos de una familia invariante frente a la acción de un grupo y no tenemos otro conocimiento previo sobre el modelo estadístico paramétrico a analizar (esto es, no nos encontramos en una situación donde fuera recomendable el uso de métodos Bayesianos), la solución a nuestro problema debería reflejar la propiedad de invariancia de la familia. En otras palabras, la clase de estimadores razonables sobre los que finalmente seleccionar, si es posible, uno óptimo, debe ser la de los equivariantes y no será conveniente buscar un estimador en una clase más amplia, ya que esto puede dar lugar a situaciones paradójicas o incluso contradictorias.

Un ejemplo destacado al respecto lo tenemos en Stein [37]. En este trabajo de 1956 se pretende estimar la media de la distribución normal m -variante $N_m(\mu, I)$ utilizando como función de pérdida:

$$L(\mu, d) = (d - \mu)' (d - \mu) , \quad d = d(X) \in \mathbb{R}^m , \quad X \in \mathbb{R}^m$$

Es inmediato verificar que el estimador máximo verosímil para la media μ es $\hat{d} = X$. Además, este estimador es insesgado y su función de riesgo asociada a la pérdida anterior es:

$$R(\hat{d}, \mu) = m ; \quad \forall \mu \in \mathbb{R}^m$$

Aunque \hat{d} ha sido considerado como un estimador óptimo en todos los sentidos (en lo que a cuestiones de invariancia se refiere, es equivariante frente a la acción del grupo aditivo de \mathbb{R}^m) el autor prueba que el estimador máximo verosímil es admisible si $m \leq 2$ pero inadmisibles si $m \geq 3$. En tal caso, el estimador:

$$\tilde{d} = \left(1 - \frac{m-2}{X'X}\right) X$$

es preferible, en términos de riesgo, a \hat{d} pues:

$$R(\tilde{d}, \mu) = m - (m-2) E_{\mu} \left(\frac{1}{m-2+2K} \right), \quad \forall \mu \in \mathbb{R}^m$$

donde $K \sim \text{Poiss}(\frac{1}{2}\mu'\mu)$. Sin embargo, \tilde{d} no es recomendable para estimar μ pues no refleja la invariancia de la familia $N_m(\mu, I)$ frente a traslaciones. Esta falta de invariancia se traduce en que cambios de unidades en las variables o reparametrizaciones del modelo llevarían a estimaciones contradictorias de \tilde{d} . Más detalles al respecto pueden ser consultados en Muirhead [30] o bien en el propio trabajo de Stein.

Problemas similares se presentan al estimar la matriz de varianzas covarianzas de una distribución normal multivariante. Supongamos que disponemos de N vectores aleatorios independientes X_1, \dots, X_N que estén idénticamente distribuidos según una ley normal $N_m(\mu, \Sigma)$. Sea:

$$A = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})'$$

Es conocido que el estimador máximo verosímil para Σ es $\hat{\Sigma} = \frac{1}{N}A$. Sin embargo, este estimador es sesgado con lo que habitualmente se trabaja con una corrección insesgada de $\hat{\Sigma}$, esto es con la matriz de varianzas covarianzas muestral $S = \frac{1}{n}A$, siendo $n = N - 1$.

Siguiendo con esta idea de proponer correcciones del estimador máximo verosímil para estimar Σ , James y Stein [38] en su trabajo de 1961 están interesados en un estimador de la matriz de varianzas covarianzas que sea función de A . Utilizan como pérdida:

$$L(\phi, \Sigma) = \text{tr}(\Sigma^{-1}\phi) - \log \det(\Sigma^{-1}\phi) - m \quad (1)$$

donde ϕ es ahora una matriz $m \times m$ simétrica y definida no negativa.

Dada la naturaleza de la familia Normal, parece razonable exigir que si $\phi(A)$ es un estimador de Σ y L es una matriz $m \times m$ no singular, entonces ϕ debe verificar:

$$\phi(M'AM) = M'\phi(A)M \quad (2)$$

La justificación queda clara si recordamos que para una familia invariante una transformación en el espacio muestral revierte en una transformación en el espacio de parámetros. Así, $\phi(M'AM)$ debe estimar $M'\Sigma M$ de igual manera que $M'\phi(A)M$ lo hace.

Si imponemos la condición (2) para todas las matrices L no podemos obtener otra información que $\phi(A) = \alpha A$ donde α es una constante a determinar. Considerando como función de pérdida L , definida en (1), se verifica que el estimador de mínimo riesgo para Σ de la forma αA es la matriz de varianzas-covarianzas muestral S .

James y Stein, en el trabajo de 1961 al que nos hemos referido anteriormente, proponen una variante de este procedimiento y buscan el estimador de mínimo riesgo para Σ que sea función de A entre todos aquellos que verifican la condición (2) para las matrices triangulares superiores. Denotemoslo por $\phi^*(A)$. No daremos la forma explícita de este estimador para m arbitraria, pues para nuestros propósitos basta con examinar el resultado obtenido cuando $m = 2$. Escribamos la matriz A como $(a_{ij})_{i,j=1,2}$. Entonces:

$$\phi^*(A) = \begin{pmatrix} \frac{a_{11}}{n+1} & \frac{a_{12}}{n+1} \\ \frac{a_{21}}{n+1} & \frac{a_{22}}{n-1} - \frac{2a_{12}^2}{(n^2-1)a_{11}} \end{pmatrix}$$

Aunque en términos de riesgo el estimador $\phi^*(A)$ sea preferible entre otros a S , debe tenerse en cuenta que no es invariante frente a permutaciones de las variables, condición esta última que lo hace incoherente para estimar a Σ . No es extraño que hayamos obtenido este resultado, pues una permutación de las variables se puede expresar como una transformación de éstas vía una matriz ortogonal que toma la forma, para $m = 2$:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Al exigir la condición (2) únicamente para las matrices triangulares superiores no estamos considerando todas las posibles transformaciones lineales a las que se puede someter a Σ y, entre las que estamos descartando, se encuentra la de las permutaciones de las variables.

Pese a lo que hemos visto en las dos situaciones anteriores, no todas las familias son invariantes frente a la acción de algún grupo. Brillinger [8] proporciona condiciones necesarias y suficientes para que un problema estadístico sea invariante bajo la acción de un grupo. Habitualmente dispondremos de familias que no serán invariantes bajo la acción de un grupo pero para las que es posible que su transformación por éste no quede *demasiado alejada*

de un desplazamiento de la familia original, por ejemplo si las transformaciones que hemos considerado son *casi* la identidad. En esta situación de *casi* invariancia, ¿debemos desestimar toda la información que proporciona la teoría clásica? ¿Tenemos que escoger entre el blanco y el negro?

Este trabajo pretende ser una contribución a la estimación puntual donde se analice el problema de selección de un estimador para los grises que van desde la existencia de invariancia clásica hasta la ausencia total de ésta. Sus diferentes *tonos* serán los órdenes de invariancia que vamos a definir.

El contexto donde se van a desarrollar todas estas ideas es el de la Geometría Diferencial Informativa. Ésta es una de las aplicaciones de la Geometría Diferencial en Estadística. La idea de considerar la matriz de Información de Fisher como medida local de la distancia entre miembros de una familia de distribuciones de probabilidad, llevó a Rao [34] en 1945 a introducir una distancia Riemanniana en el espacio de parámetros. La distancia obtenida puede ser considerada como una generalización natural de la distancia de Mahalanobis [29]. Más recientemente, Burbea y Rao [11] en 1982, Burbea [9] en 1986 y Oller y Burbea [10] en 1988 han contribuido, entre muchos otros, con sus trabajos al desarrollo de la Geometría Diferencial Informativa. Amari y Barndorff-Nielsen también deben ser citados en esta introducción pues sus aportaciones a la Geometría Diferencial en Estadística son más que relevantes. Pese a que la mayor parte de su trabajo quizá no esté conectado directamente con el material que se aquí se presenta, en Amari [2] y Amari, Barndorff-Nielsen *et al.* [3] pueden consultarse temas relacionados.

El primer capítulo de esta memoria debe entenderse como una aportación más a la hora de ilustrar la conveniencia de incorporar la invariancia en los procedimientos de selección de estimadores. En muchas situaciones los problemas estadísticos son demasiado *amplios*, en el sentido de que encontrar una solución concreta para un problema puede resultar una tarea ardua. Habitualmente se emplea un procedimiento de reducción de problema por suficiencia: trabajar espacio de los estadísticos suficientes minimales conlleva una reducción de la dimensión del espacio muestral sin que esto afecte a la información relevante, en lo que a la estimación de los parámetros se refiere, contenida en la muestra.

En nuestra situación, sabemos que la familia Normal univariante es invariante bajo la acción del grupo afín real. La reducción por invariancia permite la restricción a una clase razonable de estimadores: aquellos que sean equivariantes bajo la acción del grupo afín. Dentro de esta familia probamos la existencia y unicidad de un elemento para el que el Riesgo Riemanniano (la función de pérdida que se considera es la distancia de Rao al cuadrado) es mínimo. Este estimador es *óptimo*, por lo menos en el sentido que acabaríamos de introducir: uniformemente de mínimo riesgo entre todos los equivariantes.

La pregunta es, ¿qué ocurre cuando no hay invariancia? ¿Podremos reproducir un proceso similar en la situación de *casi* invariancia a la que nos hemos referido previamente?

Antes de continuar, tenemos que sustituir el término *casi* invariancia por *invariancia en orden*. Éste no será definido hasta el capítulo 3 pero continuar utilizándolo en esta introducción podría llevar a confusión: el concepto de casi invariancia ya tiene un significado concreto dentro de la Estadística Matemática tal y como puede consultarse en cualquier referencia clásica sobre el tema.

Supongamos la existencia de una familia de distribuciones de probabilidad en un espacio medible (Ω, \mathcal{A}) y de una familia de transformaciones biyectivas y bimedibles en el espacio muestral. En primer lugar, y aceptando que ha existido una perturbación de las probabilidades debida a estas transformaciones, necesitábamos saber cuan lejos se encontraba la nueva familia de la original.

Este requerimiento hizo necesario demostrar, bajo ciertas condiciones, un teorema de *existencia de proyección* para perturbaciones infinitesimales, el teorema 2.2.2, según una medida que hemos llamado de *discrepancia*. Debemos observar que no podíamos utilizar el teorema de la proyección clásico pues no necesariamente estábamos en un contexto Hilbertiano. Este resultado puede ser calificado como razonable o natural y sorprender la inclusión de una demostración detallada en este trabajo. De hecho, Amari [2] utiliza un resultado similar al trabajar con α -divergencias, pero sin proporcionar una demostración detallada. En cualquier caso, en los textos clásicos de Geometría Diferencial que han sido consultados (Lang [26], O’Neil [32], [12], Hicks [23], Petersen [33], entre muchos otros) no se ha podido localizar una demostración que supiera a la que aquí se proporciona.

La sección 2.3 está dedicada a probar que las hipótesis del teorema 2.2.2 se cumplen variedades Riemannianas de dimensión finita y en el espacio de Hilbert de las funciones de cuadrado integrable si seleccionamos en el primero la distancia de Rao al cuadrado como discrepancia y en el segundo la de Hellinger al cuadrado. Estos dos entornos serán en los que habitualmente nos encontraremos.

Una vez establecida la existencia de esta proyección es posible pasar a la definición de familia invariante en orden, tomando como punto de mira las invariantes clásicas. Tal y como se verá en el capítulo 3, éstas últimas quedan englobadas en las familias invariantes en orden como las de orden $+\infty$.

En los capítulos 4 y 5 introducimos los conceptos de función de pérdida invariante en orden y estimador equivariante en orden. De esta manera quedarían estudiados, en orden, los tres elementos básicos de la invariancia clásica: familia, pérdida y estimador. Obtenemos que las definiciones en orden engloban

a las clásicas como las de orden $+\infty$ y que como éstas permiten la obtención de algunos resultados destacados. Entre los que se detallan en los capítulos 4 y 5 deben destacarse dos.

En primer lugar, la distancia de Rao al cuadrado es una función de pérdida invariante de orden, como mínimo, el de la familia. De esta manera reproducimos, en orden, el conocido resultado que establece la invariancia de la distancia de Rao al cuadrado frente a la acción de un grupo.

Por otra parte, y dentro de los estimadores equivariantes en orden, debe hacerse notar que, para estos, el riesgo y el sesgo son cantidades controladas en un orden que depende del orden de la familia. Hemos obtenido así una clase *razonable* de estimadores, pues los equivariantes en orden, además de recoger las propiedades de invariancia de la familia, mantienen su sesgo y su riesgo limitado. Son conocidas las conexiones entre riesgo, sesgo e invariancia establecidas por Lehmann [27] en 1951. El capítulo 6 está dedicado a intentar establecer estas mismas conexiones pero ahora convenientemente relajadas por el contexto de la invariancia en orden.

Al explorar todas estas cuestiones observamos que nuestros análisis se basaban en la comparación de medidas de probabilidad: las originales frente a las transformadas y las proyectadas en la familia de partida. El origen de la perturbación no era relevante, simplemente su existencia en el espacio de las probabilidades. Esto nos ha llevado a considerar *la estabilidad en orden* de familias, pérdidas y estimadores. Ésta engloba la invariancia en orden de igual manera que la invariancia clásica puede entenderse como la de orden $+\infty$. El capítulo 7 es una breve introducción al respecto de estas ideas, que esperamos se vean completadas en un futuro próximo con nuevos resultados.

Índice General

Introducción	i
1 Familia normal univariante	1
1.1 Estimadores equivariantes	1
1.2 Estimador m.i.r.e.	3
1.3 Evaluación numérica	5
1.4 Familia $N(0, \sigma^2)$	10
2 Existencia de proyección local	15
2.1 Discrepancia. Proyección	15
2.2 Proyección local según una discrepancia	16
2.3 Dos ejemplos destacados	23
3 Familia invariante en orden	43
3.1 Perturbaciones en el espacio muestral	43
3.2 Familia invariante de orden α_I	48
3.3 Variedad ambiente	52
3.4 Ejemplos.	56
3.5 Transformación inducida inversa aproximada.	59
4 Pérdida invariante en orden	63
4.1 Pérdida invariante de orden β_{SC} y β_V	63
4.2 Distancia de Rao	67
4.3 Invariancia de orden β_W y β_S de ρ^2	70
4.4 Ejemplos.	84
5 Estimador equivariante en orden	85
5.1 Estimador equivariante de órdenes γ_W y γ_S	85
5.2 Equivariancia en orden, riesgo y sesgo	88
5.3 Ejemplos	95

6	Lehmann y la invariancia en orden	101
6.1	Estimadores insesgados en sentido clásico	101
6.2	Orden de conmutatividad	102
6.3	Estimadores insesgados en una dirección	106
7	Estimación estable	113
7.1	Familias localmente 1-paramétricas	113
7.2	Familias y pérdidas estables	119
7.3	Estimadores estables	121
8	Conclusiones	145
8.1	Resumen de resultados de la memoria	145
8.2	Perspectivas de la Estimación Estable	146
A	Geometría Diferencial	149
A.1	Variedades	149
A.2	Vectores tangentes. Espacio tangente.	151
A.3	Aplicación Jacobiana. Hessiano.	152
A.4	Conexiones	154
A.5	Aplicación exponencial.	156
A.6	Subvariedades	156
A.7	Variedades Riemannianas	157
A.8	Difeomorfismos. Isometrías.	161
A.9	Campos de Jacobi	162
A.10	Campos de Killing	171

Índice de Figuras

1.1	Representación gráfica de la evaluación numérica y de la aproximación sugerida de λ_n , $\lambda_n^* = 1 + \frac{4}{n}$	6
1.2	Representación conjunta de la ganancia, en términos de riesgo intrínseco, al usar el estimador equivariante de mínimo riesgo intrínseco y su aproximación sugerida en lugar del estimador máximo verosímil.	9
1.3	Representación conjunta de la ganancia, en términos de sesgo intrínseco, al usar el estimador equivariante de mínimo riesgo intrínseco y su aproximación sugerida en lugar del estimador máximo verosímil.	9
2.1	Representación gráfica que ilustra la no unicidad de la proyección para el centro de la circunferencia.	17
2.2	Muchos puntos en el eje de simetría no admiten proyección única sobre la parábola.	18
2.3	Ningún punto del eje de simetría de la figura puede ser proyectado sobre Θ	18
3.1	Representación gráfica de la no invertibilidad, en general, de las transformaciones $\theta \mapsto \overline{g}_t \theta$	48
4.1	Representación gráfica conjunta de $\rho_{\Theta}^2(\overline{g}_t \theta_1, \overline{g}_t \theta_2)$ y $\rho_{\Theta}^2(\theta_1, \theta_2)$ frente a $\tau_{\theta_1}^{\overline{g}_t \theta_1} \exp_{\overline{g}_t \theta_1}^{-1} \overline{g}_t \theta_2$ y $\exp_{\theta_1}^{-1} \theta_2$	73
A.1	Subconjunto inmerso en \mathbb{R}^2 que no es un subespacio topológico de \mathbb{R}^2	157

Capítulo 1

Familia normal univariante

Este primer capítulo está dividido en tres bloques. En primer lugar se caracteriza la clase de los estimadores equivariantes bajo la acción del grupo afín para el modelo normal univariante. Un segundo punto consiste en demostrar la existencia y unicidad de un estimador, perteneciente a la clase de los equivariantes, de mínimo Riesgo intrínseco Riemanniano, esto es, con mínimo valor medio de la distancia de Rao al cuadrado. Denotaremos este estimador como estimador m.i.r.e. Finalmente, y debido a que por el momento no se ha encontrado una expresión analítica para el estimador m.i.r.e., se ha evaluado numéricamente. Los resultados obtenidos se encuentran en la sección 1.3 de este capítulo.

1.1 Estimadores equivariantes

Consideremos el modelo normal univariante $\mathcal{P}_\Theta = \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \Theta\}$ donde Θ , el espacio de parámetros, es la variedad 2-dimensional $\mathbb{R} \times \mathbb{R}_+$ y $N(\mu, \sigma^2)$ denota la distribución normal con media μ y varianza σ^2 .

Observemos que esta familia de distribuciones \mathcal{P}_Θ es invariante bajo la acción del grupo afín $\mathcal{G} = \{g_{a,b} \mid (a,b) \in \mathbb{R}^2, a \neq 0\}$, donde la acción viene definida por:

$$g_{a,b}(x) = ax + b, \quad x \in \mathbb{R}$$

Así, tal y como hemos comentado en la introducción de este capítulo, para encontrar el estimador de mínimo riesgo Riemanniano vamos a restringir nuestra atención a la clase de los estimadores equivariantes de los parámetros. Por otra parte, dada una muestra de tamaño n y sin pérdida de generalidad, podemos buscar el estimador intrínsecamente óptimo entre todos los estimadores que sean función del estadístico suficiente minimal (\overline{X}_n, S_n^2) donde \overline{X}_n y S_n^2

son la media y la varianza basadas en n observaciones independientes. Una justificación detallada al respecto de estos temas debe ser consultada en Giri [20], Wijsman [39] o en Muirhead [30] entre otros.

En este punto podemos plantearnos la siguiente cuestión. Estamos abordando la demostración de la existencia y unicidad del estimador m.i.r.e. reduciendo primero el problema por invariancia y una vez aquí imponiendo que el estimador buscado sea función del estadístico suficiente. Si hubiéramos alterado el orden entre invariancia y suficiencia, ¿el resultado final hubiera sido el mismo?. En general no, pero debido a que los elementos de \mathcal{P}_Θ pueden ser factorizados de manera conveniente, los resultados de Hall, Wijsman y Gosh [21] son válidos en nuestro contexto y, por tanto, la reducción del problema primero por suficiencia y luego por invariancia es equivalente a exigir primero invariancia y a continuación realizar la reducción por suficiencia.

Con estas ideas en mente, observemos que el grupo afín induce de manera natural una acción en el espacio muestral del estadístico suficiente minimal dada por $\bar{g}_{a,b}(\bar{x}, s^2) = (a\bar{x} + b, a^2 s^2)$, $(\bar{x}, s^2) \in \mathbb{R} \times \mathbb{R}_+$. Cabe destacar además que esta acción coincide con la del grupo inducido sobre el espacio de parámetros y ambas serán denotadas como $\bar{g}_{a,b}$.

Podemos ahora caracterizar los estimadores equivariantes que sean función de (\bar{X}_n, S_n^2) según la siguiente proposición.

Proposición 1.1.1. *Con las notaciones anteriores, sea $U : \mathbb{R} \times \mathbb{R}_+ \rightarrow \Theta$, un estimador equivariante de (μ, σ^2) , esto es, un estimador tal que:*

$$U(\bar{g}_{a,b}(\bar{x}, s^2)) = \bar{g}_{a,b}U(\bar{x}, s^2)$$

Entonces, $U \in \{U_\lambda, \lambda > 0\}$ donde:

$$U_\lambda(\bar{x}, s^2) = (\bar{x}, \lambda s^2), \quad \lambda > 0. \quad (1.1)$$

Demostración. Escribamos $U = (U_1, U_2)$. La condición de equivariancia para U implica que:

$$U_1(a\bar{x} + b, a^2 s^2) = a U_1(\bar{x}, s^2) + b, \quad U_2(a\bar{x} + b, a^2 s^2) = a^2 U_2(\bar{x}, s^2)$$

donde estas igualdades se verifican para todo $\bar{x} \in \mathbb{R}$ y $s^2 > 0$. Tomando (\bar{x}, s^2) igual a $(0, 1)$ obtenemos:

$$U_1(b, a^2) = a U_1(0, 1) + b, \quad U_2(b, a^2) = a^2 U_2(0, 1)$$

Por otra parte, observemos que:

$$U_1(b, a^2) = a U_1(0, 1) + b \quad \text{y} \quad U_1(b, (-a)^2) = -a U_1(0, 1) + b$$

Así, $U_1(0, 1)$ debe ser 0. Si ahora recopilamos toda la información, obtenemos que $U_1(\bar{x}, s^2) = \bar{x}$ y $U_2(\bar{x}, s^2) = \lambda s^2$ tal y como queríamos ver. □

1.2 Estimador m.i.r.e.

En esta sección probaremos que para el modelo normal univariante y bajo la acción del grupo afín existe, y además es único, un estimador equivariante de mínimo riesgo. Para el riesgo utilizaremos como función de pérdida la distancia Riemanniana al cuadrado.

La matriz de información de Fisher de una familia regular de distribuciones de probabilidad induce una estructura Riemanniana en el espacio de parámetros, ver Rao [34]. La distancia correspondiente, también conocida como la *distancia de Rao*, puede entenderse como una generalización de la distancia de Mahalanobis y su cuadrado utilizarse como función de pérdida natural. Más detalles sobre este tema deben ser consultados en Oller y Corcuera [31].

La proposición 1.1.1 prueba que los estimadores equivariantes para el modelo normal univariante vienen dados por (1.1), una familia que contiene el estimador máximo verosímil. Gracias a este resultado, podemos pasar ahora a seleccionar el *mejor* estimador equivariante en el contexto del análisis intrínseco, esto es el estimador que minimice el riesgo tomando como función de pérdida la distancia de Rao al cuadrado.

La distancia de Rao ρ entre los puntos $(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2) \in \Theta$ viene dada por:

$$\rho((\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)) = \sqrt{2} \log \frac{1 + \delta(1, 2)}{1 - \delta(1, 2)}$$

donde:

$$\delta(1, 2) = \left(\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2} \right)^{1/2}$$

Observemos que δ queda inalterada por transformaciones afines del modelo normal univariante. Esta propiedad implica que ρ^2 define una pérdida invariante bajo la acción de $\overline{\mathcal{G}}$. Además, como $\overline{\mathcal{G}}$ actúa transitivamente en Θ , el riesgo de cualquier estimador equivariante es constante sobre todo el espacio de parámetros Θ . De esta manera, es suficiente minimizar el riesgo en un punto de Θ , por ejemplo en $(0, 1) \in \Theta$.

Sin embargo, antes de continuar debemos verificar que el riesgo Riemanniano del estimador U_λ tiene sentido para todo $\lambda > 0$. Denotemos la esperanza respecto a la distribución normal $N(\mu, \sigma^2)$ por $E_{(\mu, \sigma^2)}$ y sea $E \equiv E_{(0, 1)}$.

Proposición 1.2.1. *Si $n \geq 3$:*

$$E_{(\mu, \sigma^2)} (\rho^2((\bar{x}, \lambda s^2), (\mu, \sigma^2))) = E (\rho^2((\bar{x}, \lambda s^2), (0, 1))) < \infty, \forall \lambda > 0$$

para cualquier $(\mu, \sigma^2) \in \Theta$.

Demostración. Recordemos que el grupo $\overline{\mathcal{G}}$ actúa transitivamente en Θ . Además U_λ es equivariante y ρ^2 es invariante bajo la acción de \mathcal{G} . De esta manera, el riesgo Riemanniano debe ser constante y basta probar, tal y como se ha comentado anteriormente, que es finito en $(0, 1) \in \Theta$.

Para el modelo normal univariante, una distancia de Mahalanobis generalizada, ver Burbea y Oller [10], viene dada por:

$$d_{MG}((\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)) = \frac{|\mu_1 - \mu_2|}{(\sigma_1 \sigma_2)^{\frac{1}{2}}}$$

y verifica:

$$\rho^2((\bar{x}, \lambda s^2), (0, 1)) \leq d_{MG}^2((\bar{x}, \lambda s^2), (0, 1)) + 2 \frac{(\sqrt{\lambda s^2} - 1)^2}{\sqrt{\lambda s^2}}. \quad (1.2)$$

Si ahora evaluamos la esperanza, debido a (1.2) y para $n \geq 3$, obtenemos que:

$$E(\rho^2((\bar{x}, \lambda s^2), (0, 1))) < +\infty, \quad \forall \lambda > 0$$

con lo que la proposición queda demostrada. □

El resultado anterior nos permite introducir:

$$\Phi_n(\lambda) = E(\rho^2((\overline{X}_n, \lambda S_n^2), (0, 1))), \quad \lambda > 0, \quad (1.3)$$

esto es el riesgo Riemanniano de U_λ . Vamos a demostrar que existe un único elemento λ_n para el que se alcanza el mínimo de la aplicación Φ_n .

Para obtener la existencia de λ_n necesitamos utilizar el siguiente resultado de Geometría Diferencial que puede ser consultado en Chavel [12], pp. 336–338, entre otros.

Teorema 1.2.2. *Sea \mathcal{N} un conjunto convexo en una variedad Riemanniana geodésicamente completa \mathcal{M} . Si la curvatura seccional en cada punto de \mathcal{N} es a lo sumo 0 o una constante $\mathcal{K} > 0$ tal que $\text{diam}(\mathcal{N}) < \frac{\pi}{2\sqrt{\mathcal{K}}}$, entonces el cuadrado de la distancia Riemanniana es una función estrictamente convexa.*

Detalles adicionales respecto a conjuntos convexos y a funciones sobre variedades pueden ser consultados Karcher [24] y Kobayashi y Nomizu [25].

En el modelo normal univariante la curvatura Riemanniana (Gaussiana) es constante e igual a $-\frac{1}{2}$. Estamos, por lo tanto, en las hipótesis del teorema 1.2.2 y podemos afirmar que si $t \mapsto c(t)$ es una geodésica, la función real $t \mapsto \rho^2(c(t), \theta)$ es estrictamente convexa, donde $\theta \in \Theta$ es arbitrario.

Por otra parte, observemos que módulo reparametrizaciones la curva $\gamma(\lambda) = (\bar{x}, \lambda s^2)$ es una geodésica en Θ . En efecto: sea $\tau = \frac{\sqrt{2}}{2} \log \lambda$ el parámetro arco. Entonces, la curva $\tilde{\gamma}(\tau) = \gamma(e^{\sqrt{2}\tau}) = (\bar{x}, e^{\sqrt{2}\tau} s^2)$ cumple las ecuaciones de las geodésicas de la métrica informativa para el modelo normal univariante, ver Atkinson y Mitchell [5] o Burbea y Rao [11], y su grafo coincide con el de $\gamma(\lambda)$. De esta manera, a la curva $\tau \mapsto \tilde{\gamma}(\tau)$ sí que podemos aplicarle el teorema anterior y afirmar la aplicación $\tau \mapsto \rho^2(\tilde{\gamma}(\tau), (0, 1))$ es estrictamente convexa.

Sea $F_n(\tau) \equiv \Phi_n(e^{\sqrt{2}\tau})$ el riesgo Riemanniano en $\lambda = e^{\sqrt{2}\tau}$. Como la convexidad estricta de un proceso estocástico es heredada por la media del proceso, podemos asegurar que la aplicación $\tau \mapsto F_n(\tau)$ es también estrictamente convexa. Por otra parte, y gracias al Lema de Fatou, $F_n(\tau) \rightarrow +\infty$ si $\tau \rightarrow 0$ o $\tau \rightarrow \infty$. Esto implica, junto con la convexidad estricta de F_n , que existe un único elemento τ_n en el que la función F_n alcanza su valor mínimo.

Finalmente, como la aplicación $\tau \mapsto e^{\sqrt{2}\tau}$ es una función creciente, también debe existir un único λ_n tal que $\Phi_n(\lambda_n) = \min_{\lambda>0} \Phi_n(\lambda)$. Queda así demostrada la existencia y unicidad, que habíamos anunciado, del estimador m.i.r.e. para la familia normal univariante bajo la acción del grupo afín.

1.3 Evaluación numérica

Debido a que no ha sido posible encontrar una expresión analítica para λ_n , se ha optado por realizar una evaluación numérica del mismo. Ésta se ha resuelto utilizando métodos de Monte Carlo y por medio de un programa escrito en lenguaje Fortran, el cual para diferentes valores del tamaño muestral, n , evalúa numéricamente λ_n .

El valor λ_n se ha obtenido evaluando la raíz de Φ'_n por el método de la Regula-Falsi y con una tolerancia igual a 10^{-8} . Para ello, en primer lugar, se ha tenido que aproximar numéricamente el riesgo Riemanniano:

$$\Phi_n(\lambda) = E(\rho^2((\bar{X}_n, \lambda S_n^2), (0, 1)))$$

Esto se ha conseguido mediante la simulación de una distribución $N(0, \frac{1}{n})$ y de una χ^2_{n-1} independientes. Entonces, como este riesgo es una esperanza, la doble integral puede ser aproximada vía Ley Fuerte de los Grandes Números, la cual garantiza que la evaluación empírica converge casi seguramente al valor real, cuando n aumenta.

Los resultados numéricos obtenidos por este proceso junto a una aproximación sencilla de λ_n , que denotaremos $\lambda_n^* \equiv 1 + 4/n$, se encuentran en la

tabla 1.1. Las evaluaciones numéricas de $\Phi_n(\lambda_n)$, $\Phi_n(1)$ y de $\Phi_n(\lambda_n^*)$ también se muestran en la tabla 1.1. Si representamos conjuntamente λ_n y su aproximación λ_n^* obtenemos:

Parámetro óptimo

2

1

Evaluación de la aproximación numérica de λ_n
Aproximación sugerida para λ_n

10 20 30 40 50 60 70 80 90 100

Figura 1.1: Representación gráfica de la evaluación numérica y de la aproximación sugerida de λ_n , $\lambda_n^* = 1 + \frac{4}{n}$.

Observemos que el comportamiento de $U_{\lambda_n^*}$, en términos del Riesgo Riemanniano, es similar al de U_{λ_n} , el estimador equivariante de mínimo Riesgo Intrínseco Riemanniano (m.i.r.e.). Más aún, para tamaños muestrales pequeños, $n < 10$, la reducción del riesgo, respecto al estimador máximo verosímil (m.l.e.) se encuentra entre el 10 y el 25%, siendo claramente preferible el estimador m.i.r.e. o su aproximación. Para tamaños muestrales grandes, esto es $n > 30$, la reducción del riesgo es menor que 5%, y por tanto el m.l.e. puede ser usado en lugar del m.i.r.e. por simplicidad. Los detalles pueden consultarse en la tabla 1.1 y en la figura 1.2.

Otro aspecto es el sesgo intrínseco del estimador obtenido. De hecho, conexiones entre mínimo riesgo, sesgo e invariancia han sido establecidas, ver Lehmann [27]. Sin embargo, como en nuestra situación la acción del grupo no es conmutativa, no podemos asegurar que el estimador m.i.r.e. sea insesgado y por tanto debemos realizar una análisis independiente.

Consideremos $\theta \in \Theta$ y $\lambda > 0$ y sea $B_n(\lambda, \theta) = E_\theta(\exp_\theta^{-1}(\overline{X}_n, \lambda S_n^2))$, el tensor de sesgo de $(\overline{X}_n, \lambda S_n^2)$ donde \exp_θ es la aplicación exponencial correspondiente a la conexión de Levi-Civita evaluada en $\theta \in \Theta$. Recordemos que la definición de sesgo generalizado puede ser consultada en Lehmann [27] y cuestiones correspondientes al análisis intrínseco en Oller y Corcuera [31].

Siguiendo las ideas desarrolladas en la sección 1.2 para el riesgo intrínseco Riemanniano, se demuestra que $\|B_n(\lambda, \theta)\|^2$ (donde la norma se evalúa en θ) también es invariante bajo transformaciones afines. Además, como la acción del grupo es transitiva $\|B_n(\lambda, \theta)\|^2$ debe ser constante sobre el espacio de parámetros. Denotaremos $\|B_n(\lambda, (0, 1))\|^2$ como $\|B_n(\lambda)\|^2$. La tabla 1.2

muestra los resultados numéricos para $\|B_n(\lambda_n)\|^2$, $\|B_n(1)\|^2$ y $\|B_n(\lambda_n^*)\|^2$. Observemos que la reducción del sesgo respecto al estimador m.l., en términos de la norma al cuadrado del sesgo, se encuentra entre el 20 y el 5% para tamaños muestrales pequeños, $n < 10$, siendo preferible el estimador m.i.r.e.. Para tamaños muestrales grandes, $n > 20$, la reducción de la norma al cuadrado del sesgo es menor que el 5%. Así vemos que el comportamiento de $U_{\lambda_n^*}$, en términos del sesgo intrínseco, no es tan bueno como en términos del riesgo intrínseco como puede observarse en la tabla 1.2 o bien en la representación gráfica 1.3.

Tabla 1.1: Resultados numéricos para λ_n , λ_n^* , $\Phi_n(\lambda_n)$, $\Phi_n(1)$ y $\Phi_n(\lambda_n^*)$.

n	4	5	6	7	8	9	10	20
λ_n	2.158	1.796	1.605	1.487	1.408	1.351	1.308	1.138
λ_n^*	2.000	1.800	1.667	1.571	1.500	1.444	1.400	1.200
$\Phi_n(\lambda_n)$	0.719	0.522	0.411	0.339	0.289	0.252	0.223	0.105
$\Phi_n(1)$	1.025	0.698	0.526	0.420	0.349	0.298	0.260	0.114
$\Phi_n(\lambda_n^*)$	0.722	0.522	0.411	0.341	0.291	0.254	0.226	0.106

n	30	40	50	60	70	80	90	100
λ_n	1.089	1.065	1.052	1.042	1.036	1.032	1.028	1.025
λ_n^*	1.133	1.100	1.080	1.066	1.057	1.050	1.044	1.040
$\Phi_n(\lambda_n)$	0.069	0.051	0.040	0.033	0.028	0.025	0.022	0.020
$\Phi_n(1)$	0.072	0.053	0.042	0.034	0.029	0.025	0.022	0.020
$\Phi_n(\lambda_n^*)$	0.069	0.051	0.041	0.034	0.029	0.025	0.022	0.020

Tabla 1.2: Resultados numéricos para $\|B_n(\lambda_n)\|^2$, $\|B_n(1)\|^2$ y $\|B_n(\lambda_n^*)\|^2$.

n	4	5	6	7	8	9	10	20
$\ B_n(\lambda_n)\ ^2$	0.438	0.309	0.238	0.194	0.163	0.140	0.123	0.055
$\ B_n(1)\ ^2$	0.518	0.354	0.267	0.213	0.177	0.151	0.131	0.057
$\ B_n(\lambda_n^*)\ ^2$	0.422	0.310	0.245	0.202	0.172	0.149	0.132	0.059

n	30	40	50	60	70	80	90	100
$\ B_n(\lambda_n)\ ^2$	0.035	0.026	0.020	0.017	0.014	0.012	0.011	0.010
$\ B_n(1)\ ^2$	0.036	0.027	0.021	0.017	0.014	0.012	0.011	0.010
$\ B_n(\lambda_n^*)\ ^2$	0.037	0.027	0.021	0.017	0.015	0.013	0.011	0.010

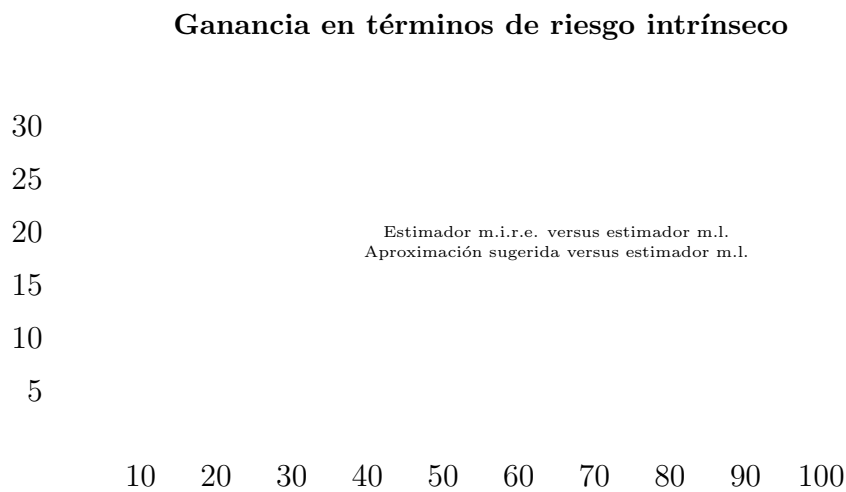


Figura 1.2: Representación conjunta de la ganancia, en términos de riesgo intrínseco, al usar el estimador equivariante de mínimo riesgo intrínseco y su aproximación sugerida en lugar del estimador máximo verosímil.

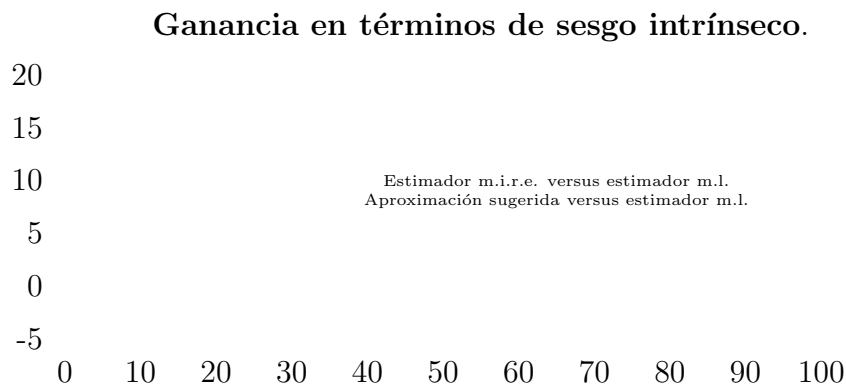


Figura 1.3: Representación conjunta de la ganancia, en términos de sesgo intrínseco, al usar el estimador equivariante de mínimo riesgo intrínseco y su aproximación sugerida en lugar del estimador máximo-verosímil.

1.4 Familia $N(0, \sigma^2)$

Como punto final a este capítulo, consideremos la familia de distribuciones de probabilidad:

$$\mathcal{Q}_\Theta = \{N(0, \sigma^2), \sigma^2 \in \Theta\}$$

donde ahora Θ , el espacio de parámetros, es la variedad 1-dimensional \mathbb{R}_+ . El grupo multiplicativo real:

$$g_a(x) = a x, \quad x \in \mathbb{R}, \quad a \in \mathbb{R}$$

actúa sobre \mathcal{Q}_Θ pues transforma una variable aleatoria X , con ley $N(0, \sigma^2)$, en la variable $g_a X$, que sigue una distribución $N(0, a^2 \sigma^2)$. Denotemos como \mathcal{G} el grupo de transformaciones $\{g_a | a \in \mathbb{R}\}$.

Así, estamos en una situación similar a la de la familia \mathcal{P}_Θ de los apartados anteriores (aquí Θ indica $\mathbb{R} \times \mathbb{R}_+$) pues disponemos de una familia, \mathcal{Q}_Θ , que es invariante bajo la acción de un grupo. Nuestro objetivo es reproducir, en la medida de lo posible, los razonamientos de las secciones 1.1 y 1.2 pero para la familia que ahora nos ocupa: queremos encontrar un estimador equivariante y de mínimo riesgo Riemanniano para σ^2 .

Nuevamente reduciremos el problema por suficiencia. Es posible verificar que las condiciones de factorización gracias a las que la reducción del problema por suficiencia y luego por invariancia es equivalente a la reducción por invariancia y más tarde por suficiencia (ver Hall, Wijsman y Ghosh [21]) también se dan en la familia \mathcal{Q}_Θ . Por lo tanto, de igual manera a como procedimos en la sección 1.1, podemos pasar a seleccionar los estimadores equivariantes como función de S_n^2 , donde S_n^2 indica la varianza basada en n observaciones independientes.

Antes de continuar, observemos que el grupo multiplicativo real induce en el espacio del estadístico suficiente minimal S_n^2 una acción que viene dada por $\bar{g}_a s^2 = a^2 s^2$, $s^2 \in \mathbb{R}_+$. Esta acción coincide con la existente en el espacio de parámetros y denotaremos ambas como \bar{g}_a .

El siguiente resultado caracteriza los estimadores equivariantes para σ^2 bajo la acción de \mathcal{G} .

Proposición 1.4.1. *Con las notaciones precedentes, sea $U : \mathbb{R}_+ \rightarrow \Theta$ un estimador equivariante de σ^2 bajo la acción del grupo multiplicativo real, esto es, un estimador tal que:*

$$U(\bar{g}_a s^2) = \bar{g}_a U(s^2)$$

Entonces, $U \in \{U_\lambda, \lambda > 0\}$ donde:

$$U_\lambda(s^2) = \lambda s^2, \quad \lambda > 0$$

Demostración. Por ser U equivariante bajo la acción de \mathcal{G} , se debe verificar:

$$U(a^2 s^2) = a^2 U(s^2), \quad a^2 \in \mathbb{R}_+, \quad s^2 \in \mathbb{R}_+$$

Pero la igualdad anterior debe ser cierta para cualquier $s^2 > 0$. Tomando $s^2 = 1$, obtenemos:

$$U(a^2) = a^2 U(1), \quad a \in \mathbb{R}_+$$

Observemos que $U(1) > 0$ pues $U : \mathbb{R}_+ \rightarrow \Theta$. Así, si denotamos $U(1)$ como λ , la proposición queda demostrada. \square

El resultado anterior prueba que cada elemento de la familia de estimadores para σ^2 :

$$\{\lambda S_n^2, \quad \lambda > 0\}$$

es equivariante bajo la acción de \mathcal{G} . Nuestro objetivo es seleccionar un valor de λ óptimo, en el sentido que minimice el riesgo tomando como función de pérdida la distancia Riemanniana al cuadrado.

Vamos a verificar, en primer lugar, que nuestro criterio de selección de λ tiene sentido. Sabemos (ver Burbea [9]) que la distancia de Rao al cuadrado, que denotaremos ρ^2 , entre dos elementos $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$ de la familia \mathcal{Q}_Θ viene dada por:

$$\rho^2(\sigma_1^2, \sigma_2^2) = 2 \log^2\left(\frac{\sigma_1}{\sigma_2}\right)$$

Tal y como observamos en la sección (1.2) de este capítulo, basta verificar que en un punto del espacio de parámetros, por ejemplo en $\sigma^2 = 1$, la esperanza de la distancia de Rao al cuadrado entre λS_n^2 y σ^2 es finita. Denotemos por E_{σ^2} la esperanza respecto a la distribución $N(0, \sigma^2)$ y sea $E \equiv E_1$. Se verifica que:

$$\begin{aligned} E\left(\rho^2(\lambda S_n^2, 1)\right) &= E\left(2 \log^2\left(\frac{\lambda S_n^2}{1}\right)\right) = 2 E\left(\log^2\left(\frac{2\lambda}{n} \frac{\chi_{n-1}^2}{2}\right)\right) = \\ &= 2 \left(\log^2\left(\frac{2\lambda}{n}\right) + 2 \log\left(\frac{2\lambda}{n}\right) E\left(\log\left(\frac{\chi_{n-1}^2}{2}\right)\right) + E\left(\log^2\left(\frac{\chi_{n-1}^2}{2}\right)\right) \right) \end{aligned}$$

Sea $k \geq 1$. Estudiemos a parte:

$$\begin{aligned} E \left(\log^k \left(\frac{\chi_{n-1}^2}{2} \right) \right) &= \int_0^\infty \log^k \left(\frac{v}{2} \right) \frac{2^{-\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} v^{\frac{n-1}{2}-1} e^{-\frac{1}{2}v} dv = \\ &= \frac{1}{\Gamma(\frac{n-1}{2})} \int_0^\infty \log^k(t) t^{\frac{n-1}{2}-1} e^{-t} dt = \frac{\Gamma^k(\frac{n-1}{2})}{\Gamma(\frac{n-1}{2})} \end{aligned}$$

donde Γ es la función Gamma de Euler. De esta manera:

$$E \left(\rho^2(\lambda S_n^2, 1) \right) = 2 \left(\log^2 \left(\frac{2\lambda}{n} \right) + 2 \log \left(\frac{2\lambda}{n} \right) \frac{\Gamma'(\frac{n-1}{2})}{\Gamma(\frac{n-1}{2})} + \frac{\Gamma^2(\frac{n-1}{2})}{\Gamma(\frac{n-1}{2})} \right)$$

Es posible verificar que en nuestro caso se dan las condiciones para poder derivar bajo el signo integral. Así:

$$\frac{d}{d\lambda} E \left(\rho^2(\lambda S_n^2, 1) \right) = \frac{4}{\lambda} \left(\log \left(\frac{2\lambda}{n} \right) + \Psi \left(\frac{n-1}{2} \right) \right)$$

siendo Ψ la función Digamma. Si recordamos que $\lambda > 0$ e igualamos a cero obtenemos que el valor crítico, λ_n , buscado es:

$$\lambda_n = \frac{n}{2} e^{-\Psi(\frac{n-1}{2})} \quad (1.4)$$

Para estudiar el comportamiento de λ_n , al aumentar n , recurriremos a una fórmula asintótica para la función Digamma que puede encontrarse en la sección 6.4.11 de Abramovitz, [1]:

$$\Psi \left(\frac{n-1}{2} \right) \sim -\log \left(\frac{n-1}{2} \right) + \frac{1}{n-1} + \sum_{k=1}^{\infty} \frac{B_{2k}}{2k} \frac{2^k}{(n-1)^k}$$

donde B_n son los números de Bernoulli. Observemos que esta expresión prueba que $\lim_{n \rightarrow \infty} \lambda_n = 1$. La evaluación numérica de λ_n a partir de la expresión (1.4) así lo confirma.

Con este ejemplo finalizamos el primer capítulo de este trabajo, que se ha dedicado a estudiar algunas aplicaciones prácticas de la invariancia. Sin embargo, antes debemos realizar una observación. Se podría pensar en un tercer ejemplo enmarcado en este ámbito y que sería el de la familia $N(\mu, 1)$ bajo la acción del grupo aditivo real. En este caso, el único estimador equivariante función del estadístico minimal $\overline{X_n}$ (donde recordemos que $\overline{X_n}$ es la media basada en n observaciones independientes) es el propio $\overline{X_n}$. No tendría sentido realizar un estudio similar a los que hemos presentado pues no existe el problema de seleccionar un estimador *óptimo* entre los de una clase o familia.

Tabla 1.3: Resultados numéricos para λ_n .

n	1	2	3	4	5	6	7	8	9	10
λ_n	3.562	1.781	1.446	1.310	1.237	1.192	1.161	1.139	1.122	1.110

n	20	30	40	50	60	70	80	90	100
λ_n	1.052	1.034	1.025	1.020	1.017	1.014	1.013	1.011	1.010

En este capítulo hemos estudiado con detalle la importancia de la propiedad de invariancia en la elección de una regla de decisión para un problema estimación puntual. El hecho de que la familia sea invariante ha simplificado considerablemente esta selección tal y como hemos podido verificar en la sección 1.1 y en esta última. Más detalles al respecto también pueden ser consultados en García y Oller [18].

En los siguientes capítulos pretendemos ocuparnos de modelos paramétricos y de familias de transformaciones para las que *casi*, término que será concretado más adelante, se tenga una situación de invariancia e incorporar, cuando sea necesario, esta propiedad en los diferentes procedimientos de Inferencia Estadística.

Capítulo 2

Existencia de proyección local

Este capítulo está dividido en tres partes. En la primera se introduce el concepto de proyección local respecto a lo que definiremos como una *discrepancia*. En la segunda se prueba que, bajo ciertas condiciones, elementos de una variedad Δ pueden ser proyectados localmente, y de manera única, sobre una subvariedad de Δ . Tal y como veremos en el siguiente capítulo, este resultado permitirá definir una familia de transformaciones en el espacio de parámetros que desempeñará un papel similar al del grupo inducido en la situación de invariancia clásica. La tercera sección de este capítulo está dedicada a analizar el problema de la proyección en dos ejemplos destacados.

2.1 Discrepancia. Proyección

Sea Δ una variedad de clase \mathcal{C}^∞ . Es un resultado conocido, que recordaremos en el apéndice correspondiente a Geometría Diferencial, que el producto cartesiano de variedades también tiene estructura de variedad. De esta manera, podremos considerar aplicaciones de esta variedad en cualquier otra a las que además tendrá sentido exigirles tanta suavidad como la situación requiera. De momento sea δ una aplicación:

$$\begin{aligned}\delta : \Delta \times \Delta &\longrightarrow [0, +\infty) \\ (p, \theta) &\longmapsto \delta(p, \theta)\end{aligned}$$

tal que $\delta(p, \theta) = 0 \Leftrightarrow p = \theta$.

Definición 2.1.1. Diremos que δ es una *discrepancia* en $\Delta \times \Delta$.

Una discrepancia debe entenderse como una medida de no-proximidad entre los elementos de Δ . Para nuestros ulteriores propósitos Δ va a ser una

variedad \mathcal{C}^∞ en la que sus puntos son *etiquetas* de una familia paramétrica de distribuciones de probabilidad. Teniendo en cuenta este contexto, es interesante observar que la aplicación δ no tiene porque ser una distancia entre probabilidades o una f -divergencia aunque éstos serían dos ejemplos típicos de discrepancias.

Sea ahora $p_0 \in \Delta$ y $\Theta \subset \Delta$ una subvariedad de Δ . Introducimos:

$$\delta(p_0, \Theta) := \inf_{\theta \in \Theta} \delta(p_0, \theta)$$

Estamos interesados en analizar si existen, y bajo qué condiciones, elementos de la subvariedad para los que se alcance este ínfimo.

Definición 2.1.2. Diremos que θ_0 es una *proyección según δ de p_0 en Θ* si y sólo si $\theta_0 \in \Theta$ y $\delta(p_0, \Theta) = \delta(p_0, \theta_0)$.

Por simplicidad de notación, nos referiremos a los puntos proyección según una discrepancia como puntos proyección, si la medida de discrepancia queda clara por el contexto.

Observemos, finalmente, que la restricción de δ a $\Delta \times \Theta$ induce de forma natural una medida de discrepancia en este conjunto. Este va a ser nuestro ámbito de trabajo, pues no estaremos interesados en discrepancias entre pares $(p, q) \in \Delta \times \Delta$ sino entre pares $(p, \theta) \in \Delta \times \Theta$ tal y como veremos en la siguiente sección.

2.2 Proyección local según una discrepancia

En primer lugar, recordemos el teorema de la proyección clásico en espacios unitarios cuya demostración se puede encontrar en Rudin [36] o en Balkrishnan [6] entre muchos otros.

Teorema 2.2.1. *Sea $(E, (\cdot, \cdot)_E)$ espacio unitario.*

1. *Sea C un subconjunto convexo completo de E . Entonces, $\forall x \in E$, $\exists!$ $y \in C$ aproximación óptima (según la distancia definida a través de $(\cdot, \cdot)_E$) de x sobre C .*
2. *Sea F un subespacio vectorial completo de E . Entonces, $E = F \oplus F^\perp$*

Nuestro objetivo es reproducir, en la medida de lo posible, el primer apartado del teorema anterior. Es de suponer que no podremos obtener un resultado similar pues las hipótesis que exigimos son bastante más débiles.

La existencia de proyección local sobre una subvariedad Θ de una variedad Δ de clase \mathcal{C}^∞ no es un resultado tan directo como pueda pensarse a priori. Recordemos que $\Theta \subset \Delta$ es una *subvariedad* de Δ si y sólo si:

1. Θ es un subespacio topológico de Δ .
2. La inclusión $i : \Theta \hookrightarrow \Delta$ es de clase \mathcal{C}^∞ y la aplicación diferencial $(i_*)_p$ es $1 - 1$.

Además, suponiendo que la proyección exista, no tiene porqué ser, en general, única tal y como se muestra en los siguientes ejemplos.

Ejemplo 2.2.1. Sea Δ el espacio euclídeo 2-dimensional, \mathbb{R}^2 . Consideremos la subvariedad de Δ dada por \mathcal{S}^1 y tomemos como discrepancia la distancia Euclídea al cuadrado. Para cualquier punto del plano euclídeo, la proyección sobre \mathcal{S}^1 existe y es única salvo para el centro de la circunferencia que, tal y como se muestra en la figura, admite infinitas proyecciones.

Figura 2.1: Representación gráfica que ilustra la no unicidad de la proyección para el centro de la circunferencia.

En el siguiente ejemplo veremos que podemos tener más de un punto problemático a la hora de proyectar.

Ejemplo 2.2.2. Consideremos, de nuevo en \mathbb{R}^2 , la subvariedad dada por la parábola Θ y representada en la figura 2.2 dada por (x, x^2) , $x \in \mathbb{R}$. Tomemos como discrepancia la distancia Euclídea al cuadrado. En el eje de simetría de la parábola podemos encontrar puntos conflictivos pues su proyección sobre ésta no es única. Sin embargo, observemos que a medida que nos acerquemos a la subvariedad, y aún estando sobre el eje de simetría, podremos proyectar localmente y de forma única sobre Θ .

Por otra parte, observemos que con la definición de subvariedad adoptada (ver O’Neil [32] o Lang [26] para más detalles al respecto) se excluyen subconjuntos de \mathbb{R}^2 como el que mostramos en el siguiente ejemplo.



$$\Theta$$

Figura 2.2: Muchos puntos en el eje de simetría no admiten proyección única sobre la parábola.

Ejemplo 2.2.3. Continuando con \mathbb{R}^2 , consideremos el subespacio topológico Θ y tomemos una vez más como discrepancia la distancia Euclídea al cuadrado. En la representación gráfica de la figura 2.3 obtenemos nuevamente dificultades en el eje de simetría de la figura a la hora de proyectar que, en este caso, no conseguiríamos solucionar ni acercándonos a Θ .



$$\theta$$

Figura 2.3: Ningún punto del eje de simetría de la figura puede ser proyectado sobre Θ .

Los ejemplos anteriores, en particular el ejemplo 2.2.2, parecen indicar que la suavidad de la subvariedad en un entorno de un punto $\theta \in \Theta$ es un factor crucial a la hora de proyectar de manera única puntos suficientemente próximos a Θ . En el ejemplo 2.2.3 la necesidad de la suavidad, queda de manifiesto: esta condición implicaría que pudiéramos aproximar localmente Θ por planos tangentes a la subvariedad y que por tanto la situación se transforme en aproximadamente Euclídea. Claramente en el vértice $\theta \in \Theta$ del ejemplo 2.2.3 no lo podemos hacer.

Pasamos ahora a enunciar el resultado principal de esta sección y para el que el Teorema de la Función Implícita en espacios de Banach, no necesariamente

de dimensión finita, va a ser una herramienta clave. Este teorema puede encontrarse en cualquier texto básico de Análisis Funcional aunque en Lang [26] también se encuentra demostrado.

Teorema 2.2.2. *Sea Δ una E -variedad de clase \mathcal{C}^∞ y $\Theta \subset \Delta$ una subvariedad de Δ , modelada en un espacio de Banach F .*

Sea δ una medida de discrepancia en Δ de clase \mathcal{C}^r ($r \geq 2$) y $p_0 \in \Delta$ cumpliendo:

i. $\exists! \theta_0 \in \Theta$ tal que $\delta(p_0, \Theta) = \delta(p_0, \theta_0)$.

ii. $Hess^{\delta(p_0, \cdot), \theta_0}$ es definido positivo.

Entonces, existen U_0 entorno abierto de p_0 en Δ , V_0 entorno de θ_0 en Θ y una aplicación $\Upsilon : U_0 \longrightarrow V_0$ unívocamente determinada y también de clase \mathcal{C}^r de manera que:

$$\forall p \in U_0 : \quad \delta(p, \Upsilon(p)) = \delta(p, \Theta)$$

Demostración. Sea (U, ϕ) una carta en p_0 y (V, ψ) una carta en θ_0 . Tal y como puede encontrarse en el Apéndice correspondiente a Geometría Diferencial de este trabajo o con más detalle en Lang [26], $\forall \theta \in V$ el espacio tangente $T_\theta(\Theta)$ y F están en biyección por medio de la aplicación:

$$\begin{aligned} J_\theta : F &\longrightarrow T_\theta(\Theta) \\ v &\longmapsto (V, \psi, v) \end{aligned}$$

con el convenio de notar la clase de equivalencia $[(V, \psi, v)]$ a través de su representante (V, ψ, v) . Esta biyección dota a $T_\theta(\Theta)$ de estructura de espacio vectorial topológico, demostrándose que ésta es independiente de la carta (V, ψ) seleccionada.

Sea $\mathcal{L}(F, \mathbb{R})$ el dual topológico de F y consideremos la aplicación:

$$\begin{aligned} \bar{g} : \phi(U) \times \psi(V) &\longrightarrow \mathcal{L}(F, \mathbb{R}) \\ (a, b) &\longmapsto g(\phi^{-1}(a), \psi^{-1}(b)) \end{aligned}$$

donde $\forall (p, \theta) \in U \times V$:

$$\begin{aligned} g(p, \theta) : F &\longrightarrow T_\theta(\Theta) \longrightarrow \mathbb{R} \\ v &\longmapsto J_\theta(v) \longmapsto D\delta(p, \psi^{-1}(\cdot))|_{\psi(\theta)}(v) \end{aligned}$$

siendo D la diferencial ordinaria de una aplicación de F en \mathbb{R} y recordando que $J_\theta(v) = (V, \psi, v)$. Observemos que, fijados $p \in U$ y $\theta \in V$, $g(p, \theta)$ es efectivamente un elemento de $\mathcal{L}(F, \mathbb{R})$ pues la biyección entre F y $T_\theta(\Theta)$ es lineal por construcción.

Sabemos que $\delta(p_0, \theta_0) = \delta(p_0, \Theta)$. Entonces, la aplicación $\delta(p_0, \cdot)$ tiene un mínimo en θ_0 con lo que $g(p_0, \theta_0) = 0$ y $\bar{g}(\phi(p_0), \psi(\theta_0)) = 0$.

Vamos a demostrar que, además:

$$D_2 \bar{g}(\phi(p_0), \psi(\theta_0)) : F \longrightarrow \mathcal{L}(F, \mathbb{R})$$

es un isomorfismo topológico, donde D_2 indica la derivada parcial respecto al segundo argumento de la aplicación \bar{g} . De esta manera, podremos utilizar el Teorema de la Función Implícita entre espacios de Banach y a través de las cartas en p_0 y en θ_0 obtener resultados en la variedad.

Se verifica que:

$$D_2 \bar{g}(\phi(p_0), \psi(\theta_0)) = D g(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}$$

Por otra parte:

$$g(p_0, \cdot)_{*, \theta_0} : T_{\theta_0}(\Theta) \longrightarrow \mathcal{L}(F, \mathbb{R})$$

cumple que si $\bar{v} = (V, \psi, v)$ es un elemento de $T_{\theta_0}(\Theta)$, $g(p_0, \cdot)_{*, \theta_0}(\bar{v})$ es, por definición de jacobiano, el elemento de $\mathcal{L}(F, \mathbb{R})$ dado por:

$$D g(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}(v)$$

Estamos utilizando en este punto la identificación, por medio de la carta identidad, de la variedad de Banach $\mathcal{L}(F, \mathbb{R})$ modelada sobre ella misma con el espacio tangente en cualquiera de sus puntos.

Así, para demostrar que $D_2 \bar{g}(\phi(p_0), \psi(\theta_0))$ es un isomorfismo topológico, basta que probemos que $g(p_0, \cdot)_{*, \theta_0}$ lo es.

Examinemos con más detalle el elemento:

$$D g(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}(v) \in \mathcal{L}(F, \mathbb{R}), \quad v \in F$$

Sea $w \in F$. Entonces:

$$D g(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}(v)(w) = D^2 \delta(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}(v, w)$$

Denotaremos $D^2 \delta(p_0, \psi^{-1}(\cdot))|_{\psi(\theta_0)}(v, w)$, como $B(v, w)$, $\forall v, w \in F$. De esta manera, y por medio de una carta en θ_0 , $g(p_0, \cdot)_{*, \theta_0}$ se identifica con la aplicación:

$$\begin{aligned} F &\longrightarrow \mathcal{L}(F, \mathbb{R}) \\ w &\longrightarrow B(\cdot, w) \end{aligned}$$

Recordemos que nuestro objetivo era demostrar que la aplicación anterior es un isomorfismo topológico. La linealidad está clara pues la diferencial segunda es bilineal y por tanto lineal en cada una de sus componentes. Además, δ es de clase \mathcal{C}^r siendo $r \geq 2$, con lo que la continuidad también queda establecida.

Probemos ahora la inyectividad. Adaptaremos la demostración del teorema de dualidad de Fisher–Riesz a nuestro caso particular. La demostración de este teorema puede encontrarse en cualquier texto sobre Análisis Funcional, como por ejemplo [36]. Sea $w \in F$ tal que $B(\cdot, w) = 0 \in \mathcal{L}(F, \mathbb{R})$. Queremos ver que $w = 0$. En efecto:

$$B(\cdot, w) = 0 \Leftrightarrow \forall v \in F : B(v, w) = 0 \Leftrightarrow w = 0$$

pues B es definida positiva por hipótesis.

Finalmente probemos la exhaustividad. Sea $\alpha \in \mathcal{L}(F, \mathbb{R})$. Queremos ver que existe $w \in F$ tal que $\alpha(\cdot) = B(\cdot, w)$. Si $\alpha = 0$, basta tomar $w = 0$. Por lo tanto, podemos suponer que $\alpha \neq 0$.

Introducimos $F_\alpha = \alpha^{-1}(\{0\})$. F_α es un subespacio de F ya que $F_\alpha = \ker(\alpha)$. Además es cerrado pues $\{0\}$ es cerrado de \mathbb{R} y α es continua. Esto implica, por ser F completo, que F_α también es completo. Por otra parte, $B(\cdot, \cdot)$ puede entenderse como un producto escalar en F con lo que el par $(F, B(\cdot, \cdot))$ es un espacio unitario. Podemos aplicar la segunda parte del teorema 2.2.1 y escribir:

$$F = F_\alpha \oplus F_\alpha^\perp$$

siendo $F_\alpha^\perp \neq \emptyset$ ya que $\alpha \neq 0$ y por tanto $F_\alpha \subsetneq F$. Podremos escoger entonces $u \in F_\alpha^\perp$ tal que $B(u, u) = 1$ y observar que $\forall v \in F$, el elemento $\alpha(v) u - \alpha(u) v \in F_\alpha$. En efecto.

$$\alpha(\alpha(v) u - \alpha(u) v) = \alpha(v) \alpha(u) - \alpha(u) \alpha(v) = 0$$

De esta manera:

$$0 = B(\alpha(v) u - \alpha(u) v, u) = \alpha(v) - \alpha(u) B(v, u)$$

por ser $B(u, u) = 1$. Así:

$$\alpha(v) = \alpha(u) \quad B(v, u) = B(v, \alpha(u) \cdot u) \quad \forall v \in F$$

esto es:

$$\alpha(\cdot) = B(\cdot, \alpha(u) \cdot u)$$

con lo que la exhaustividad queda demostrada.

Podemos ahora aplicar el teorema de la Función Implícita entre espacios de Banach y concluir que:

$$\exists \bar{U} \text{ entorno abierto de } \phi(p_0) \text{ en } E$$

$$\exists \bar{V} \text{ entorno de } \psi(\theta_0) \text{ en } F$$

$$\exists \bar{\Upsilon} : \bar{U} \longrightarrow \bar{V}, \text{ unívocamente determinada y de clase } \mathcal{C}^r$$

tales que $\forall a \in \bar{U}, \bar{g}(a, \bar{\Upsilon}(a)) = 0$. Si ahora tomamos:

$$U_0 = \phi^{-1}(\bar{U} \cap \phi(U))$$

$$V_0 = \psi^{-1}(\bar{\Upsilon}(\bar{U} \cap \phi(U)) \cap \psi(V))$$

$$\Upsilon = (\psi^{-1} \circ \bar{\Upsilon} \circ \phi)|_{U_0}$$

el teorema queda demostrado. En efecto. U_0 es un entorno abierto de p_0 en Δ por construcción. V_0 es un entorno de θ_0 en Θ y la aplicación Υ es de clase \mathcal{C}^r pues para las cartas $(U_0, \phi|_{U_0})$ en p_0 y $(V_0, \psi|_{V_0})$ en θ_0 se cumple que:

$$\psi|_{V_0} \circ \Upsilon \circ \phi|_{U_0}^{-1} = \bar{\Upsilon}|_{\bar{U} \cap \phi(U)}$$

lo es.

□

Observemos que en el teorema anterior la única estructura que hemos requerido al espacio F es la de Banach. De esta manera, si el espacio sobre el que modelamos, F , es de Hilbert el resultado continua siendo válido trivialmente. En el enunciado del teorema anterior, el par (p_0, θ_0) se supone conocido. Debemos plantearnos si existen tales pares en cada situación concreta. A continuación veremos que en algunos ejemplos interesantes, como son las variedades Riemannianas de dimensión finita y el espacio de Hilbert de las funciones de cuadrado integrable, las hipótesis del teorema se verifican si nos acercamos a la subvariedad de manera conveniente y que vamos a detallar en la siguiente sección.

2.3 Dos ejemplos destacados

Este trabajo se enmarca en la perspectiva de la Geometría Diferencial Informativa, donde los modelos estadísticos paramétricos son representados mediante variedades Riemannianas que, en la mayoría de los casos que consideraremos, serán de dimensión finita. Por otra parte, si el modelo está dominado por una medida σ -finita de referencia, estas variedades están inmersas de forma natural en el espacio de Hilbert de las funciones de cuadrado integrable cuando identificamos las probabilidades con dos veces la raíz cuadrada positiva de sus densidades. Estas consideraciones nos llevan a estar interesados en analizar los resultados de la sección anterior para variedades Riemannianas de dimensión finita y para el espacio de Hilbert que hemos mencionado anteriormente. En cada uno de estos dos casos, y con las notaciones de la sección anterior, debemos verificar que las hipótesis del teorema 2.2.2 se cumplen, esto es, que existen pares de puntos $(p_0, \theta_0) \in \Delta \times \Theta$ tales que θ_0 es el único elemento de Θ para el que $\delta(p_0, \Theta) = \delta(p_0, \theta_0)$ y $Hess^{\delta(p_0, \cdot), \theta_0}$ es definido positivo.

Supongamos, en primer lugar que (Δ, ρ) es una variedad Riemanniana y que $\Theta \subset \Delta$ es una subvariedad de Δ de dimensión finita, n . Sea ρ_Θ la distancia en Θ inducida por ρ .

Notación 2.3.1. Sea $\theta \in \Theta$ y $R > 0$. Introducimos:

$$B(\theta, R) := \{p \in \Delta \mid \rho(\theta, p) < R\}$$

$$B_\Theta(\theta, R) := \{q \in \Theta \mid \rho_\Theta(\theta, q) < R\}$$

Para la demostración del teorema 2.3.1 va a ser necesario recordar algunas nociones de convexidad en una variedad Riemanniana (Δ, ρ) .

Definición 2.3.1. Diremos que $B \subset \Delta$ es *convexo* si y sólo si $\forall p, q \in B$ existe una única geodésica en B uniendo p y q y tal que la longitud de la geodésica es $\rho(p, q)$.

Definición 2.3.2. Para cada $p \in \Delta$ definimos *conv* p , el *radio de convexidad* de p como:

$$\text{conv } p = \sup\{R \mid B(p, r) \text{ es convexa } \forall r < R\}$$

Tal y como veremos a continuación, únicamente vamos a poder proyectar sobre la subvariedad si nos acercamos convenientemente a ella. Tomaremos un punto $\vartheta \in \Theta$ y construiremos explícitamente entornos de ϑ en Δ y en Θ de manera que la proyección de uno en otro pueda definirse. En la construcción

de estos entornos veremos que tanto la curvatura de la variedad como el segundo tensor fundamental van a tener un papel destacado. Más detalles al respecto de cualquiera de estos temas se pueden consultar en Lang [26] o bien O'Neill [32] o en el apéndice de Geometría Diferencial que se ha incluido en este trabajo.

Sea ahora $\vartheta \in \Theta$. Por ser (Δ, ρ) una variedad Riemanniana, el teorema de Whitehead adaptado al caso Riemanniano y cuya demostración puede encontrarse en la página 214 de Lang [26], prueba que existe $c^* > 0$ tal que $\forall 0 < r^* < c^*$ se cumple que $B(\vartheta, r^*)$ es convexo. Por otra parte, (Θ, ρ_Θ) también es una variedad Riemanniana pues Θ es una subvariedad Riemanniana de Δ . Nuevamente el teorema de Whitehead implica que existe $\bar{c} > 0$ tal que $\forall 0 < \bar{r} < \bar{c}$, $B_\Theta(\vartheta, \bar{r})$ es un conjunto convexo. Estas dos afirmaciones implican que, si indicamos:

$$\text{conv}_\Theta \vartheta = \sup\{R \mid B_\Theta(\vartheta, r) \text{ es convexa } \forall r < R\}$$

se cumple que:

$$\text{conv } \vartheta > 0$$

$$\text{conv}_\Theta \vartheta > 0$$

Sean $c(\vartheta) = \min\{\text{conv } \vartheta, \text{conv}_\Theta \vartheta\}$, k supremo de la curvatura Riemanniana seccional en $\overline{B(\vartheta, c(\vartheta))}$, esto es en $\{q \in \Delta \mid \rho(q, \vartheta) \leq c(\vartheta)\}$ y $\|V\|$ la norma del segundo tensor fundamental en Θ . Escojamos R^* y \bar{R} de la siguiente manera. Sea:

$$0 < \bar{R} < \min\left\{\frac{c(\vartheta)}{2}, \frac{1}{\|V\|}, \frac{\pi}{2\sqrt{k}}, \frac{1}{\sqrt{k}} \arctan\left(\frac{\sqrt{k}}{\|V\|}\right)\right\}$$

Para la elección de R^* distinguiremos casos en función de la curvatura Riemanniana seccional. Si $k = 0$ escogemos $0 < R^* < \text{conv } \vartheta$ tal que:

$$R^* < \frac{1}{\|V\|} - \bar{R}$$

Si k es una constante positiva tomamos $0 < R^* < \text{conv } \vartheta$ de manera que:

$$R^* < \frac{1}{\sqrt{k}} \min\left\{\frac{\pi}{2}, \arctan\left(\frac{\sqrt{k}}{\|V\|}\right)\right\} - \bar{R}$$

En cualquiera de los dos casos anteriores, impondremos adicionalmente que:

$$(B(\vartheta, 2R^*) \cap \Theta) \subset B_\Theta(\vartheta, \bar{R})$$

Teorema 2.3.1. *Con las notaciones anteriores, $\forall p \in B(\vartheta, R^*)$ existe un único $\theta \in B_\Theta(\vartheta, \overline{R})$ tal que:*

$$\rho^2(p, \theta) = \min_{\theta' \in \Theta} \rho^2(p, \theta')$$

Denotaremos el punto θ como $\Upsilon(p)$, $\forall p \in B(\vartheta, R^*)$.

Demostración. Sea $p \in B(\vartheta, R^*)$. Observemos que si $p \in \Theta$ el teorema es cierto trivialmente. Supongamos así que $p \in B(\vartheta, R^*) \cap \Theta^c$ y consideremos:

$$\begin{aligned} \gamma : [0, 1] &\longrightarrow B(\vartheta, R^*) \\ s &\longmapsto \gamma(s) \end{aligned}$$

una geodésica de Δ tal que:

$$(G1) \quad \gamma(0) = p$$

$$(G2) \quad \gamma(1) \in \Theta. \text{ Denotemos } \gamma(1) = \theta.$$

$$(G3) \quad \gamma'(1) \in T_\theta(\Theta)^\perp \subset T_\theta(\Delta) \text{ donde } \gamma'(1) \text{ indica el vector tangente a } \gamma \text{ en } s = 1.$$

Como convenio vamos a suponer que todas las curvas que consideraremos son segmentos de curva, definición A.4.4 del apéndice de Geometría Diferencial, y que por tanto los vectores tangentes en los extremos están bien definidos. Sea ahora:

$$\begin{aligned} \overline{\gamma} : [0, 1] &\longrightarrow B_\Theta(\vartheta, \overline{R}) \\ t &\longmapsto \overline{\gamma}(t) \end{aligned}$$

una geodésica de Θ de manera que $\overline{\gamma}(0) = \theta$ y tal que $\overline{\gamma}(1) \neq \overline{\gamma}(0)$. Consideremos la variación de γ dada por:

$$\begin{aligned} c_p : [0, 1] \times [0, 1] &\longmapsto B(\vartheta, R^*) \\ (s, t) &\longrightarrow \exp_p(s \exp_p^{-1} \overline{\gamma}(t)) \end{aligned}$$

siendo \exp_p la aplicación exponencial en el punto p correspondiente a la conexión de Levi-Civita, que denotaremos ∇ , en Δ . Dispondremos así de una familia de geodésicas en Δ indexadas según $t \in [0, 1]$, dadas por:

$$\begin{aligned} c_p(\cdot, t) : [0, 1] &\longmapsto B(\vartheta, R^*) \\ s &\longrightarrow c_p(s, t) \end{aligned}$$

y cumpliendo que $c_p(0, t) = p$, $c_p(1, t) = \bar{\gamma}(t)$ para todo $t \in [0, 1]$. Denotaremos $\gamma_t(s) \equiv c_p(s, t)$, $\forall (s, t) \in [0, 1] \times [0, 1]$ y observemos que $\gamma_0 = \gamma$. Si indicamos por l la longitud de una curva, vamos a demostrar que $\forall t \in (0, 1]$:

$$l^2(\gamma_t) > l^2(\gamma) \quad (2.1)$$

y que (2.1) se verifica independientemente de la geodésica $\bar{\gamma}$ escogida. De esta manera, habremos conseguido probar que:

$$\rho^2(p, \theta) = \min_{\theta' \in B_\Theta(\vartheta, \bar{R})} \rho^2(p, \theta')$$

y que tal θ es único en $B_\Theta(\vartheta, \bar{R})$. En efecto. Si hubiera otro punto $\theta' \in B_\Theta(\vartheta, \bar{R})$ para el que $\rho^2(p, \theta') \leq \min_{\tilde{\theta} \in B_\Theta(\vartheta, \bar{R})} \rho^2(p, \tilde{\theta})$, podríamos considerar, pues la bola $B_\Theta(\vartheta, \bar{R})$ convexa, una geodésica de Θ que denotaremos $\bar{\gamma}$ de manera que $\bar{\gamma}(0) = \theta$ y $\bar{\gamma}(1) = \theta'$. Entonces, con las notaciones precedentes se tendría que:

$$l^2(\gamma_1) \leq l^2(\gamma)$$

condición que contradice (2.1).

En realidad una vez que (2.1) quede demostrada, habremos probado que:

$$\rho^2(p, \theta) = \min_{\theta' \in \Theta} \rho^2(p, \theta')$$

pues si existiera otro punto $\theta' \in B_\Theta(\vartheta, \bar{R})^c \cap \Theta$ para el que:

$$\rho^2(p, \theta') < \min_{\tilde{\theta} \in \Theta} \rho^2(p, \tilde{\theta}) \quad (2.2)$$

entonces debe ser:

$$\rho(p, \theta') \leq \rho(p, \vartheta) < R^*$$

donde se ha utilizado la definición de mínimo y que $p \in B(\vartheta, R^*)$. Teniendo en cuenta esta última desigualdad:

$$\rho(\vartheta, \theta') \leq \rho(\vartheta, p) + \rho(p, \theta') \leq R^* + R^* = 2R^*$$

Así $\theta' \in B(\vartheta, 2R^*) \cap \Theta$. Pero por hipótesis $(B(\vartheta, 2R^*) \cap \Theta) \subset B_\Theta(\vartheta, \overline{R})$. Sin embargo hemos visto que $\forall \theta' \in B_\Theta(\vartheta, \overline{R})$:

$$\rho^2(p, \theta') > \rho^2(p, \theta)$$

Esto es una contradicción con (2.2). Por lo tanto:

$$\rho^2(p, \theta) = \min_{\theta' \in \Theta} \rho^2(p, \theta')$$

Queda así por demostrar, con las notaciones introducidas anteriormente, que $\forall t \in (0, 1]$:

$$l^2(\gamma_t) > l^2(\gamma)$$

Observemos que $l^2(\gamma) = l^2(\gamma_0)$. De esta manera basta demostrar que la aplicación a valores en los reales positivos y de variable real:

$$\begin{aligned} l^2 : [0, 1] &\longrightarrow [0, \infty) \\ t &\longmapsto l^2(\gamma_t) \end{aligned}$$

cumple:

$$(L1) \quad \frac{d}{dt} l^2(\gamma_t)|_{t=0} = 0$$

$$(L2) \quad l^2 \text{ es estrictamente convexa.}$$

Introducimos:

$$\begin{aligned} c'_p &= (c_p)_* \left(\frac{\partial}{\partial s} \right) \\ \dot{c}_p &= (c_p)_* \left(\frac{\partial}{\partial t} \right) \end{aligned}$$

y sean $\frac{\nabla}{\partial s}$, $\frac{\nabla}{\partial t}$ las derivadas covariantes, correspondientes a ∇ , la conexión de Levi-Civita, a lo largo de las curvas $c_p(\cdot, t)$ y $c_p(s, \cdot)$ respectivamente. Entonces:

$$\frac{d}{dt} l^2(\gamma_t) = \frac{d}{dt} \rho^2(p, \bar{\gamma}(t)) = \frac{\partial}{\partial t} \langle c'_p(s, t), c'_p(s, t) \rangle$$

ya que $\rho^2(p, \bar{\gamma}(t)) = \|c'_p(s, t)\|^2$, sin importar el valor de s , al ser $s \mapsto c_p(s, t)$ una geodésica de Δ por construcción. Así podremos afirmar también que:

$$\frac{\partial}{\partial t} \langle c'_p(s, t), c'_p(s, t) \rangle = \frac{\partial}{\partial t} \left\{ \int_0^1 \langle c'_p(s, t), c'_p(s, t) \rangle ds \right\} =$$

$$= 2 \int_0^1 \left\langle \frac{\nabla}{\partial t} c'_p(s, t), c'_p(s, t) \right\rangle ds$$

Teniendo en cuenta que la conexión de Levi-Civita es de torsión nula, se verifica que el paréntesis de Lie $[c'_p, \dot{c}_p] = 0$. Esta condición implica que $\frac{\nabla}{\partial t} c'_p = \frac{\nabla}{\partial s} \dot{c}_p$ y por tanto:

$$2 \int_0^1 \left\langle \frac{\nabla}{\partial t} c'_p(s, t), c'_p(s, t) \right\rangle ds = 2 \int_0^1 \left\langle \frac{\nabla}{\partial s} \dot{c}_p(s, t), c'_p(s, t) \right\rangle ds$$

Pero, tal y como hemos observado anteriormente, $s \mapsto c_p(s, t)$ es una geodésica de Δ . De esta manera, $\frac{\nabla}{\partial s} c'_p = 0$ y:

$$\begin{aligned} 2 \int_0^1 \left\langle \frac{\nabla}{\partial s} \dot{c}_p(s, t), c'_p(s, t) \right\rangle ds &= 2 \int_0^1 \frac{\partial}{\partial s} \langle \dot{c}_p(s, t), c'_p(s, t) \rangle ds = \\ &= 2 \langle \dot{c}_p(1, t), c'_p(1, t) \rangle \end{aligned}$$

pues recordemos que $\forall t \in [0, 1]$ $c_p(0, t) = p$ y por lo tanto $\dot{c}_p(0, t) = 0_p \in T_p(\Delta)$.

Recopilando toda la información, hemos obtenido que:

$$\frac{d}{dt} l^2(\gamma_t) = 2 \langle \dot{c}_p(1, t), c'_p(1, t) \rangle$$

Evaluyendo en $t = 0$ y si indicamos como $\dot{\bar{\gamma}}(0)$ el vector tangente a $\bar{\gamma}$ en $t = 0$:

$$\frac{d}{dt} l^2(\gamma_t)|_{t=0} = 2 \langle \dot{c}_p(1, 0), c'_p(1, 0) \rangle = 2 \langle \dot{\bar{\gamma}}(0), \gamma'(1) \rangle = 0$$

debido a la hipótesis (G3). De esta manera, la condición (L1) es cierta. Vamos a probar ahora (L2). Volviendo a derivar:

$$\frac{d^2}{dt^2} l^2(\gamma_t) = 2 \left\langle \frac{\nabla}{\partial t} \dot{c}_p(1, t), c'_p(1, t) \right\rangle + 2 \left\langle \dot{c}_p(1, t), \frac{\nabla}{\partial t} c'_p(1, t) \right\rangle$$

Analícemos en primer lugar el término correspondiente a $\langle \frac{\nabla}{\partial t} \dot{c}_p(1, t), c'_p(1, t) \rangle$. Sea $\bar{\nabla}$ la conexión Riemanniana en inducida en Θ a través de la matriz de la métrica de Δ . Entonces:

$$\frac{\nabla}{\partial t} \dot{c}_p(1, t) = \frac{\bar{\nabla}}{\partial t} \dot{c}_p(1, t) + V(\dot{c}_p(1, t), \dot{c}_p(1, t))$$

En nuestro caso $\frac{\bar{\nabla}}{\partial t} \dot{c}_p(1, t) = \frac{\bar{\nabla}}{\partial t} \dot{\bar{\gamma}}(t) = 0$ pues $\bar{\gamma}$ es una geodésica de Θ . Además, recordemos que $\frac{\nabla}{\partial t} c'_p(1, t) = \frac{\nabla}{\partial s} \dot{c}_p(1, t)$. De esta manera:

$$\begin{aligned} \frac{d^2}{dt^2} l^2(\gamma_t) &= 2 \langle V(\dot{c}_p(1, t), \dot{c}_p(1, t)), c'_p(1, t) \rangle + \\ &\quad + 2 \langle \dot{c}_p(1, t), \frac{\nabla}{\partial s} \dot{c}_p(1, t) \rangle \end{aligned}$$

Estudiemos ahora $\langle \dot{c}_p(1, t), \frac{\nabla}{\partial s} \dot{c}_p(1, t) \rangle$. Es un resultado conocido, que puede ser consultado en el apéndice correspondiente a Geometría Diferencial o en Lang [26], Chavel [12] o bien O'Neill [32], que $\dot{c}_p(s, t)$ es un campo de Jacobi a lo largo de la geodésica $s \mapsto c_p(s, t)$. Notemos este campo como $J(s)$ y descompongámoslo en sus componentes normal, $J^{nor}(s)$, y tangencial, $J^{tan}(s)$, a lo largo de la geodésica, esto es $\langle J^{tan}(s), J^{nor}(s) \rangle = 0$, $\forall s \in [0, 1]$. Con estas notaciones:

$$\begin{aligned} \langle \dot{c}_p(1, t), \frac{\nabla}{\partial s} \dot{c}_p(1, t) \rangle &= \langle J(1), \frac{\nabla}{\partial s} J(1) \rangle = \\ &= \langle J^{tan}(1) + J^{nor}(1), \frac{\nabla}{\partial s} J^{tan}(1) + \frac{\nabla}{\partial s} J^{nor}(1) \rangle = \\ &= \langle J^{tan}(1), \frac{\nabla}{\partial s} J^{tan}(1) \rangle + \langle J^{tan}(1), \frac{\nabla}{\partial s} J^{nor}(1) \rangle + \\ &\quad + \langle J^{nor}(1), \frac{\nabla}{\partial s} J^{tan}(1) \rangle + \langle J^{nor}(1), \frac{\nabla}{\partial s} J^{nor}(1) \rangle \end{aligned}$$

Estudiaremos cada uno de los cuatro sumandos anteriores por separado. Como consecuencia de la ecuación de Jacobi, se tiene:

$$J^{tan}(s) = (a + bs)c'_p(s, t)$$

Pero $J(0) = \dot{c}_p(0, t) = 0_p \in T_p(\Delta)$ tal y como ya se ha visto. Así $a = 0$ y:

$$J^{tan}(s) = bs \, c'_p(s, t) \tag{2.3}$$

Teniendo en cuenta esta expresión:

$$\frac{\nabla}{\partial s} J^{tan}(s) = b \, c'_p(s, t) + bs \, \frac{\nabla}{\partial s} c'_p(s, t)$$

Recordemos que $\frac{\nabla}{\partial s} c'_p(s, t) = 0$. De esta manera, si evaluamos la ecuación anterior en $s = 1$ y tenemos en cuenta (2.3), también evaluada en $s = 1$, obtenemos que $\frac{\nabla}{\partial s} J^{tan}(1) = J^{tan}(1)$ con lo que:

$$\langle J^{tan}(1), \frac{\nabla}{\partial s} J^{tan}(1) \rangle = \|J(1)\|^2$$

y:

$$\langle J^{nor}(1), \frac{\nabla}{\partial s} J^{tan}(1) \rangle = 0 \quad (2.4)$$

Estudiemos ahora la parte normal de J . Utilizaremos los teoremas de comparación de Rauch. Estos teoremas pueden encontrarse demostrados en Chavel [12].

Sean Σ y σ cotas superior e inferior de la curvatura Riemanniana seccional a lo largo de la geodésica $s \mapsto c_p(s, t)$. Observemos que tanto Σ como σ pueden depender de t . Por el teorema de comparación de Rauch para campos de Jacobi, y teniendo en cuenta que la geodésica $s \mapsto c_p(s, t)$ no es unitaria, podemos escribir:

$$\begin{aligned} f_{\Sigma}(\|c'_p(1, t)\|) \langle \frac{1}{\|c'_p(1, t)\|} \frac{\nabla}{\partial s} J^{nor}(1), J^{nor}(1) \rangle &\geq \\ &\geq f'_{\Sigma}(\|c'_p(1, t)\|) \langle J^{nor}(1), J^{nor}(1) \rangle \end{aligned}$$

Así:

$$\langle J^{nor}(1), \frac{\nabla}{\partial s} J^{nor}(1) \rangle \geq \frac{f'_{\Sigma}(\|c'_p(1, t)\|)}{\|c'_p(1, t)\| f_{\Sigma}(\|c'_p(1, t)\|)} \|J(1)\|^2$$

Finalmente, recordemos que $\langle J^{tan}(s), J^{nor}(s) \rangle = 0, \forall s \in [0, 1]$. Así:

$$0 = \frac{\partial}{\partial s} \langle J^{tan}(s), J^{nor}(s) \rangle = \langle \frac{\nabla}{\partial s} J^{tan}(s), J^{nor}(s) \rangle + \langle J^{tan}(s), \frac{\nabla}{\partial s} J^{nor}(s) \rangle$$

Evalando en $s = 1$:

$$\langle J^{tan}(1), \frac{\nabla}{\partial s} J^{nor}(1) \rangle = -\langle \frac{\nabla}{\partial s} J^{tan}(1), J^{nor}(1) \rangle = 0$$

pues por (2.4), $\langle \frac{\nabla}{\partial s} J^{tan}(1), J^{nor}(1) \rangle = 0$. Recopilando toda la información obtenemos que:

$$\begin{aligned} \langle J(1), \frac{\nabla}{\partial s} J(1) \rangle &= \|J^{tan}(1)\|^2 + \langle J^{nor}(1), \frac{\nabla}{\partial s} J^{nor}(1) \rangle \geq \\ &\geq \|J^{tan}(1)\|^2 + \frac{f'_{\Sigma}(\|c'_p(1, t)\|)}{\|c'_p(1, t)\| f_{\Sigma}(\|c'_p(1, t)\|)} \|J^{nor}(1)\|^2 \end{aligned}$$

Distingamos ahora casos en función de la curvatura Riemanniana seccional. Si $\Sigma = 0$, por ser $f_{\Sigma}(0) = 0$ y al solucionar la ecuación diferencial ordinaria $f''_{\Sigma} = 0$, obtenemos que $f_{\Sigma}(t) = bt$. De esta manera:

$$\begin{aligned} \frac{d^2}{dt^2} l^2(\gamma_t) &\geq 2 \{ \|J^{tan}(1)\|^2 + \|J^{nor}(1)\|^2 + \langle V(J(1), J(1)), c'_p(1, t) \rangle \} \geq \\ &= 2 \{ \|J(1)\|^2 - \|V\| \|J(1)\|^2 \|c'_p(1, t)\| \} \end{aligned}$$

donde se ha utilizado la desigualdad de Cauchy–Schwarz. Recordemos además que $J(1) = \dot{c}_p(1, t)$ con lo que $\|J(1)\|^2 = \|\dot{c}_p(1, t)\|^2 = \rho_\Theta^2(\theta, \bar{\gamma}(1)) \neq 0$ por hipótesis. Por otra parte,

$$1 - \|V\| \|c'_p(1, t)\| = 1 - \|V\| \rho(p, \bar{\gamma}(t)) \geq 1 - \|V\| (R^* + \bar{R})$$

pues $\rho(p, \bar{\gamma}(t)) \leq \rho(p, \vartheta) + \rho(\vartheta, \bar{\gamma}(t)) \leq \rho(p, \vartheta) + \rho_\Theta(\vartheta, \bar{\gamma}(t)) < R^* + \bar{R}$. Pero R^* ha sido escogido de manera que:

$$R^* < \frac{1}{\|V\|} - \bar{R}$$

esto es $1 - \|V\| (R^* + \bar{R}) > 0$. De esta manera, hemos obtenido que, en el caso de curvaturas Riemannianas negativas o cero:

$$\frac{d^2}{dt^2} l^2(\gamma_t) \geq \rho_\Theta^2(\theta, \bar{\gamma}(1)) \{1 - \|V\| (R^* + \bar{R})\} > 0$$

esto es la aplicación $t \mapsto l^2(\gamma_t)$ es estrictamente convexa.

Para curvaturas positivas, podemos tomar $\Sigma = k$ que también será una cota superior de la curvatura a lo largo de la geodésica $s \mapsto c_p(s, t)$. En este caso $f_k(t) = \sin(\sqrt{k}t)$ con lo que:

$$\frac{f'_k(\|c'_p(1, t)\|)}{c'_p(1, t) f_k(\|c'_p(1, t)\|)} = \frac{\sqrt{k} \|c'_p(1, t)\|}{\tan(\sqrt{k} \|c'_p(1, t)\|)}$$

Pero $0 < \frac{x}{\tan(x)} < 1$ si $x \in (0, \frac{\pi}{2})$. Ese es nuestro caso, pues $\sqrt{k} \|c'_p(1, t)\| = \sqrt{k} \rho(p, \bar{\gamma}(t)) < \sqrt{k} (R^* + \bar{R}) < \frac{\pi}{2}$ por hipótesis. Así:

$$\begin{aligned} \frac{d^2}{dt^2} l^2(\gamma_t) &\geq 2 \{ \|J^{tan}(1)\|^2 + \\ &+ \frac{\sqrt{k} \|c'_p(1, t)\|}{\tan(\sqrt{k} \|c'_p(1, t)\|)} \|J^{nor}(1)\|^2 + \langle V(J(1), J(1)), c'_p(1, t) \rangle \} \geq \\ &\geq 2 \|J(1)\|^2 \|c'_p(1, t)\| \left\{ \frac{\sqrt{k}}{\tan(\sqrt{k} \|c'_p(1, t)\|)} - \|V\| \right\} \end{aligned}$$

Ya sabemos que tanto $\|J(1)\|^2$ como $\|c'_p(1, t)\|$ son valores positivos. Así, para ver (L2) basta verificar que:

$$\frac{\sqrt{k}}{\tan(\sqrt{k} \|c'_p(1, t)\|)} - \|V\| > 0$$

Pero esta condición es cierta debido a que:

$$R^* + \bar{R} < \frac{1}{\sqrt{k}} \arctan\left(\frac{\sqrt{k}}{\|V\|}\right)$$

y teniendo en cuenta que $\|c'_p(1, t)\| < R^* + \bar{R}$, para todo $t \in [0, 1]$. Se ha establecido así la convexidad estricta de la aplicación:

$$\begin{aligned} [0, 1] &: \longrightarrow [0, \infty) \\ t &\longmapsto l^2(\gamma_t) \end{aligned}$$

y el teorema queda demostrado. □

El siguiente corolario prueba que además de verificarse la existencia y la unicidad de la proyección local para puntos p convenientemente próximos a la subvariedad, el Hessiano de la aplicación $\rho^2(p, \cdot)$ evaluado en el punto $\Upsilon(p)$ es definido positivo (recordemos que $\Upsilon(p)$ se ha introducido en el enunciado del teorema 2.3.1). Podríamos aplicar así el teorema 2.2.2 tomando, con las notaciones del teorema anterior, $p_0 \in B(\vartheta, R^*)$ y $\theta_0 = \Upsilon(p_0)$ donde recordemos que $\vartheta \in \Theta$.

Corolario 2.3.2. *Con las notaciones del teorema anterior, $\forall p \in B(\vartheta, R^*)$:*

$$Hess^{\rho^2(p, \cdot), \Upsilon(p)} \text{ es definido positivo.}$$

Demostración. Sea $p \in B(\vartheta, R^*)$. Si $p \in \Theta$ que el $Hess^{\rho^2(p, \cdot), \Upsilon(p)}$ sea definido positivo es decir que $Hess^{d_N^2(f, \cdot), f}$ lo es. Pero observemos que esta condición es cierta, pues (Θ, ρ_Θ) es una variedad Riemanniana por hipótesis. Sea así $p \in B(\vartheta, R^*) \cap \Theta^c$. Consideremos:

$$\begin{aligned} \bar{\gamma}: [0, 1] &\longrightarrow B_\Theta(\vartheta, \bar{R}) \\ t &\longmapsto \bar{\gamma}(t) \end{aligned}$$

una geodésica de Θ de manera que $\bar{\gamma}(0) = \Upsilon(p)$. Hemos demostrado, en el teorema anterior que $\forall t \in [0, 1]$:

$$\frac{d^2}{dt^2} l^2(\gamma_t) > 0$$

Esto es:

$$0 < \frac{d^2}{dt^2} l^2(\gamma_t) = \sum_{i,j=1}^n \frac{\partial^2}{\partial \theta^i \partial \theta^j} \rho^2(p, \bar{\gamma}(t)) \dot{\bar{\gamma}}^i(t) \dot{\bar{\gamma}}^j(t) + \sum_{i=1}^n \frac{\partial}{\partial \theta^i} \rho^2(p, \bar{\gamma}(t)) \ddot{\bar{\gamma}}^i(t) \quad (2.5)$$

Evalutando la expresión anterior en $t = 0$, se tiene que:

$$0 < \frac{d^2}{dt^2} l^2(\gamma_t)|_{t=0} = \sum_{i,j=1}^n \frac{\partial^2}{\partial \theta^i \partial \theta^j} \rho^2(p, \Upsilon(p)) \dot{\bar{\gamma}}^i(0) \dot{\bar{\gamma}}^j(0)$$

Esto es, la matriz:

$$\left(\frac{\partial^2}{\partial \theta^i \partial \theta^j} \rho^2(p, \Upsilon(p)) \right)_{i,j=1 \dots n}$$

es definida positiva, pues (2.5) no depende de la geodésica $\bar{\gamma}$. De esta manera $Hess^{\rho^2(p, \cdot), \Upsilon(p)}$ es definido positivo y el corolario queda demostrado. \square

Tal y como se ha comentado anteriormente, también estamos interesados en la existencia de proyección local sobre una subvariedad del espacio de Hilbert H de las funciones de cuadrado integrable. En general, una subvariedad no tiene porque ser un conjunto convexo completo de H y por lo tanto no podremos recurrir al teorema 2.2.1. Debemos aplicar técnicas similares a las del teorema 2.3.1.

Sea así $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y \mathcal{Q} un conjunto de medidas finitas en (Ω, \mathcal{A}) , dominadas por una medida σ -finita μ en (Ω, \mathcal{A}) . Siguiendo las ideas de Dawid [16], la identificación de cada elemento Q de \mathcal{Q} :

$$Q \mapsto 2 \left(\frac{dQ}{d\mu} \right)^{\frac{1}{2}}$$

proporciona una inmersión global de \mathcal{Q} en $H = L^2(\Omega, \mathcal{A}, \mu)$.

Sea d la distancia definida en H a través del producto escalar ordinario. La métrica inducida en \mathcal{Q} es dos veces la *distancia de Hellinger*, que denotaremos igualmente por d . De forma explícita:

$$d(f, g) = \left(\int_{\Omega} (\sqrt{f(w)} - \sqrt{g(w)})^2 d\mu(w) \right)^{\frac{1}{2}}, \quad \forall f, g \in L^1(\Omega, \mathcal{A}, \mu), \quad f, g \geq 0$$

En nuestra situación disponemos de una familia paramétrica de distribuciones de probabilidad $\mathcal{P}_{\Theta} = \{P_{\theta} \mid \theta \in \Theta\}$ en (Ω, \mathcal{A}) de manera que $P_{\theta} \ll \mu$, $\forall \theta \in \Theta$. Sea $N = \{(\frac{dP_{\theta}}{d\mu})^{\frac{1}{2}} \mid \theta \in \Theta\}$, esto es, un subconjunto de la esfera de radio 2 de H . Exigiremos que N sea una subvariedad de H . Observemos que la distancia de Hellinger sobre N (en realidad sobre toda la esfera de radio 2) es igual a la mitad de la de Battacharyya, tal y como puede consultarse en Battacharyya [7].

Con abuso de notación, nos referiremos a la distancia de Hellinger sobre la esfera de radio 2 únicamente como la distancia de Hellinger esperando que el contexto establezca si nos referimos a la distancia o bien a la distancia inducida.

Consideremos la siguiente discrepancia en H :

$$\begin{aligned} d^2 : H \times H &\longrightarrow [0, +\infty) \\ (f, g) &\longmapsto d^2(f, g) \end{aligned}$$

Vamos a estudiar el problema de la existencia de proyección sobre N después de que la familia \mathcal{P}_{Θ} haya sido perturbada y suponiendo que las perturbaciones continúan siendo elementos de H . De momento no especificaremos la manera de perturbar y esperaremos a capítulos posteriores para discutir este tema.

Antes de pasar a analizar el problema de la proyección vamos a recordar que las geodésicas en un espacio de Hilbert son rectas.

Lema 2.3.3. *Sea $(H, (\cdot, \cdot)_H)$ un espacio de Hilbert. Entonces, las geodésicas en H son los segmentos rectilíneos.*

Demostración. Sea $c : [\alpha, \beta] \mapsto H$ una curva en H de clase \mathcal{C}^r , $r \geq 2$, tal que $c'(t) = v \in H$ y donde $c'(t)$ indica la derivación ordinaria en \mathbb{R} respecto al parámetro t . Estamos haciendo uso, en este punto de la identificación natural del espacio tangente a H en cualquier punto con el propio H . Módulo reparametrizaciones, que como es sabido no alteran la longitud de una curva, podemos suponer que $\alpha = 0$ y que $\beta = 1$.

Vamos a probar que para toda perturbación de la curva c , que denotaremos \tilde{c} , se cumple que $l(\tilde{c}) > l(c)$.

Realizaremos la demostración en dos etapas. Sea $\bar{c}(t) = c(t) + \eta(t)$ una perturbación de c de manera que $\eta'(t) = a(t)v$ para $t \in [0, 1]$. Observemos

que a debe verificar que $\int_0^1 a(t)dt = 0$ pues $\eta(0) = \eta(1) = 0$ y por tanto $0 = \eta(1) - \eta(0) = (\int_0^1 a(t)dt)v$.

Supongamos, en primer lugar, que $1 + a(t) > 0, \forall t \in [0, 1]$. Veremos que en tal caso $l(\bar{c}) = l(c)$.

$$\begin{aligned} l(\bar{c}) &= \int_0^1 \sqrt{\langle \bar{c}'(t), \bar{c}'(t) \rangle} dt = \int_0^1 \sqrt{\langle (1+a(t))v, (1+a(t))v \rangle} dt = \\ &= \left(\int_0^1 1+a(t)dt \right) \|v\| = \|v\| = l(c) \end{aligned}$$

pues $\int_0^1 a(t)dt = 0$ tal y como hemos razonado anteriormente. Este resultado no debe extrañar pues en realidad no hemos perturbado la curva c sino que hemos reparametrizado nuestro segmento rectilíneo según $s = \phi(t)$ donde $\phi'(t) = a(t), \forall t \in [0, 1]$.

Supongamos ahora que exista $t \in (0, 1)$ tal que $1 + a(t) < 0$. Por continuidad deben existir $t_0 < t_1$, puntos del intervalo $(0, 1)$ tales que $1 + a(t) > 0$ si $t \in [0, t_0]$, $1 + a(t) < 0$ si $t \in (t_0, t_1)$, $1 + a(t) > 0$ si $t \in (t_1, 1]$ y $1 + a(t_i) = 0, i = 0, 1$.

Sea $t_2 \in (t_1, 1]$ tal que $\bar{c}(t_0) = \bar{c}(t_2)$. Si indicamos por γ_α^β la restricción de una curva cualquiera γ a un intervalo de extremos α y β , se cumple que:

$$l(\bar{c}) = l(\bar{c}_0^{t_0}) + l(\bar{c}_{t_0}^{t_1}) + l(\bar{c}_{t_1}^{t_2}) + l(\bar{c}_{t_2}^1) = l(\bar{c}_0^{t_0}) + 2l(\bar{c}_{t_0}^{t_1}) + l(\bar{c}_{t_2}^1) > l(c)$$

pues $l(\bar{c}_0^{t_0}) = l(c_0^{t_0})$ y $l(\bar{c}_{t_2}^1) = l(c_{t_0}^1)$ en virtud de lo razonado si $1 + a(t) > 0$. Sea ahora $\tilde{c}(t) = c(t) + \eta(t)$, para $t \in [0, 1]$ una perturbación arbitraria de c . Descomponemos:

$$\eta(t) = \eta^{tan}(t) + \eta^{nor}(t), \quad t \in [0, 1]$$

donde $(\eta^{tan})'(t) \in \langle c'(t) \rangle$ y $(\eta^{nor})'(t) \in \langle c'(t) \rangle^\perp, \forall t \in [0, 1]$. Supondremos además que existe algún punto del intervalo $[0, 1]$ para el que la componente normal no se anula. Entonces:

$$\begin{aligned} l(\tilde{c}) &= \int_0^1 \sqrt{\langle \tilde{c}'(t), \tilde{c}'(t) \rangle} dt = \\ &= \int_0^1 \sqrt{\langle c'(t) + (\eta^{tan})'(t) + (\eta^{nor})'(t), c'(t) + (\eta^{tan})'(t) + (\eta^{nor})'(t) \rangle} dt = \\ &= \int_0^1 \sqrt{\|c'(t) + (\eta^{tan})'(t)\|^2 + \|(\eta^{nor})'(t)\|^2} dt > \int_0^1 \sqrt{\|c'(t) + (\eta^{tan})'(t)\|^2} dt = \\ &= l(c + \eta^{tan}) \end{aligned}$$

donde se ha utilizado que la existencia de un punto de $(0, 1)$ donde la componente normal no se anule implica, por continuidad nuevamente, que debe existir un entorno de este punto donde la componente normal siga sin anularse.

Pero sabemos que $l(c + \eta^{tan}) > l(c)$ con lo que queda probado que las geodésicas en un espacio de Hilbert son los segmentos rectilíneos pues son los que tienen longitud mínima.

□

Volvemos ahora al problema que nos ocupaba, esto es, la existencia de proyección respecto d_H^2 sobre una subvariedad, N , de H . Fijaremos la siguiente notación.

Notación 2.3.2. Sea $h \in H$, $R > 0$ y d_N la distancia en N inducida por d . Introducimos:

$$B(h, R) := \{f \in H \mid d(f, h) < R\}$$

$$B_N(h, R) := \{g \in N \mid d_N(g, h) < R\}$$

Para la demostración del teorema 2.3.4 utilizaremos las mismas técnicas que en el teorema 2.3.1 pero veremos que en este caso tienen un desarrollo más simple.

Observemos que en esta situación, $\forall h \in H$, $\forall r^* > 0$, $B(h, r^*)$ es un conjunto convexo. Fijemos $h \in N$ y sea:

$$conv_N h = \sup\{R \mid B_N(h, r) \text{ es convexa } \forall r < R\}$$

El teorema de Whitehead, tomando como variedad Riemanniana (N, d_N) , asegura que $conv_N h > 0$, $\forall h \in N$. Escojamos entonces:

$$0 < R^* < \frac{1}{\|V\|}$$

siendo V el segundo tensor fundamental en N , y:

$$0 < \overline{R} < \min\left\{\frac{conv_N h}{2}, \frac{1}{\|V\|} - R^*\right\}$$

Por construcción, tanto $B(h, R^*)$ como $B_N(h, \overline{R})$ son conjuntos convexos. Finalmente exigiremos tal y como lo hicimos en la situación anterior que $(B(\vartheta, 2R^*) \cap \Theta) \subset B_\Theta(\vartheta, \overline{R})$.

Teorema 2.3.4. Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de probabilidad $H = L^2(\Omega, \mathcal{A}, \mu)$ y $N \subset H$ una subvariedad de H modelada en un espacio de Banach F . Sea d la distancia en H definida a través del producto escalar ordinario en H . Entonces, con las notaciones anteriores $\forall f \in B(h, R^*)$, existe un único $g \in B_N(h, \bar{R})$ tal que:

$$d^2(f, g) = \min_{g' \in N} d^2(f, g')$$

Denotaremos $g = \Upsilon(f)$.

Demostración. Sea $f \in B(h, R^*)$. Si $f \in N$ el teorema es cierto, pues basta tomar $\Upsilon(f) = f$. Nos ocuparemos entonces de las funciones $f \in B(h, R^*) \cap N^c$. Consideremos:

$$\begin{aligned} \gamma : [0, 1] &\longrightarrow B(h, R^*) \\ s &\longmapsto \gamma(s) \end{aligned}$$

una geodésica de H de manera que:

$$(G'1) \quad \gamma(0) = f$$

$$(G'2) \quad \gamma(1) \in N. \text{ Sea } g = \gamma(1).$$

$$(G'3) \quad \gamma'(1) \in T_g(N)^\perp \subset T_g(H), \text{ siendo } \gamma'(1) \text{ el vector tangente a } \gamma \text{ en } s = 1.$$

Consideremos:

$$\begin{aligned} \bar{\gamma} : [0, 1] &\longrightarrow B_N(h, \bar{R}) \\ t &\longmapsto \bar{\gamma}(t) \end{aligned}$$

una geodésica de N tal que $\bar{\gamma}(0) = \theta$ y $\bar{\gamma}(1) \neq \bar{\gamma}(0)$.

Podemos construir una perturbación de γ , de igual manera a como lo hicimos en el teorema 2.3.1 según:

$$\begin{aligned} c_f : [0, 1] \times [0, 1] &\longmapsto B(h, R^*) \\ (s, t) &\longrightarrow \exp_f(s \exp_f^{-1} \bar{\gamma}(t)) \end{aligned}$$

donde \exp_f es la aplicación exponencial en el punto f correspondiente a la conexión de Levi-Civita, que denotaremos ∇ , en H .

Por construcción, $\forall (s, t) \in [0, 1] \times [0, 1]$ se cumple que:

$$\begin{aligned}
c_f(s, 0) &= \gamma(s) \\
c_f(1, t) &= \bar{\gamma}(s) \\
c_f(0, t) &= f
\end{aligned}$$

y $s \mapsto c_f(s, t)$ son una familia de geodésicas indexadas por $t \in [0, 1]$ que denotaremos γ_t , aunque obviamente también existe dependencia de la geodésica $\bar{\gamma}$. Sea $l^2(\gamma_t)$ la longitud al cuadrado de γ_t . Siguiendo los razonamientos detallados en el teorema 2.3.1 para demostrar que:

$$d^2(f, g) = \min_{g' \in N} d^2(f, g')$$

basta que veamos:

$$(L'1) \quad \frac{d}{dt} l^2(\gamma_t)|_{t=0} = 0$$

$$(L'2) \quad l^2 \text{ es estrictamente convexa.}$$

Sean $c'_f = (c_f)_*(\frac{\partial}{\partial s})$ y $\dot{c}_f = (c_f)_*(\frac{\partial}{\partial t})$. Tal y como hemos visto en el teorema precedente:

$$\frac{d}{dt} l^2(\gamma_t) = \frac{d}{dt} d_H^2(f, \bar{\gamma}(t)) = 2 \langle \dot{c}_f(1, t), c'_f(1, t) \rangle$$

Recordando que podemos identificar $T_f(H)$ con H pues la identidad es una carta global, en el caso que ahora nos ocupa tenemos que:

$$\begin{aligned}
\dot{c}_f &= \dot{\bar{\gamma}}(t) \\
c'_f &= \bar{\gamma}(t) - f
\end{aligned}$$

pues en un espacio de Hilbert las geodésicas son los segmentos rectilíneos tal y como se ha recordado en el lema 2.3.3. De esta manera, evaluando en $t = 0$ y debido a (G'3):

$$\frac{d}{dt} l^2(\gamma_t)|_{t=0} = 2 \{ \dot{\bar{\gamma}}(0), \bar{\gamma}(0) - f \} = 0$$

Con lo que (L'1) queda probada. Para ver (L'2), necesitamos volver a derivar la función $l^2(\gamma_t)$.

$$\frac{d^2}{dt^2} l^2(\gamma_t) = 2 \left\{ \frac{\nabla}{\partial t} \dot{\bar{\gamma}}(t), \bar{\gamma}(t) - f \right\} + 2 \{ \dot{\bar{\gamma}}(t), \dot{\bar{\gamma}}(t) \} \quad (2.6)$$

donde $\frac{\nabla}{\partial t}$ es la derivada covariante a lo largo de $\bar{\gamma}$ respecto a la conexión ∇ en H .

Se cumple que:

$$\{\dot{\bar{\gamma}}(t), \dot{\bar{\gamma}}(t)\} = d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) > 0$$

por hipótesis. Respecto al otro sumando de (2.6), si $\bar{\nabla}$ es la conexión Riemanniana inducida en N a través de la matriz de la métrica de H , se tiene que:

$$\frac{\nabla}{\partial t} \dot{\bar{\gamma}}(t) = \frac{\bar{\nabla}}{\partial t} \dot{\bar{\gamma}}(t) + V(\dot{\bar{\gamma}}(t), \dot{\bar{\gamma}}(t))$$

Pero $\bar{\gamma}$ una geodésica en N con lo que $\frac{\bar{\nabla}}{\partial t} \dot{\bar{\gamma}}(t) = 0$ y recuperando (2.6) tenemos:

$$\begin{aligned} \frac{d^2}{dt^2} l^2(\gamma_t) &= 2 \{d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) + \{V(\dot{\bar{\gamma}}(t), \dot{\bar{\gamma}}(t)), \bar{\gamma}(t) - f\}\} \geq \\ &\geq 2 \{d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) - \|V\| \|\dot{\bar{\gamma}}(t)\|^2 \|\bar{\gamma}(t) - f\|\} = \\ &= 2 d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) \{1 - \|V\| d(\bar{\gamma}(t), f)\} \end{aligned}$$

donde nuevamente se ha utilizado la desigualdad de Cauchy-Schwarz.

Pero $d(\bar{\gamma}(t), f) \leq d(\bar{\gamma}(t), h) + d(h, f) \leq d_N(\bar{\gamma}(t), h) + d(h, f) < \bar{R} + R^*$.

Así:

$$\frac{d^2}{dt^2} l^2(\gamma_t) \geq 2 d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) \{1 - \|V\| (\bar{R} + R^*)\} > 0$$

debido a las elecciones de \bar{R} y de R^* y a que $d_N^2(\bar{\gamma}(0), \bar{\gamma}(1)) > 0$ por hipótesis. Se verifica así también (L'2) y el teorema queda demostrado. \square

Finalmente y tal y como hicimos en el caso de variedades Riemannianas de dimensión finita, vamos a ver que los pares de puntos $(f, \Upsilon(f))$ que hemos obtenido en el teorema anterior verifican que $Hess^{d^2(f, \cdot), \Upsilon(f)}$ es definido positivo.

Corolario 2.3.5. *Con las notaciones del teorema anterior, $\forall f \in B(h, R^*)$, se verifica que $Hess^{d^2(f, \cdot), \Upsilon(f)}$ es definido positivo.*

Sea $f \in B(h, R^*)$. Si $f \in N$ que el $Hess^{d^2(f, \cdot), \Upsilon(f)}$ sea definido positivo es exigir que $Hess^{d_N^2(f, \cdot), f}$. Pero esta condición es cierta al ser (N, d_N) una variedad Riemanniana.

Por lo tanto, de nuevo únicamente debemos preocuparnos de los puntos $f \in B(h, R^*) \cap N^c$. Sea así:

$$\begin{aligned} \gamma : [0, 1] &\longrightarrow B(h, R^*) \\ s &\longmapsto \gamma(s) \end{aligned}$$

una geodésica de H cumpliendo las condiciones (G'1), (G'2) y (G'3) del teorema anterior y consideremos:

$$\begin{aligned} \bar{\gamma} : [0, 1] &\longrightarrow B_N(h, \bar{R}) \\ t &\longmapsto \bar{\gamma}(t) \end{aligned}$$

una geodésica de N de manera que $\bar{\gamma}(0) = \Upsilon(f)$. En el teorema 2.3.4 hemos demostrado que:

$$\frac{d^2}{dt^2} l^2(\gamma_t) > 0, \quad \forall t \in [0, 1]$$

esto es:

$$0 < \frac{d^2}{dt^2} l^2(\gamma_t) = \dot{\gamma}(t)^t D^2 d^2(f, \cdot)_{\bar{\gamma}(t)} \dot{\gamma}(t) + D d^2(f, \cdot)_{\bar{\gamma}(t)} \ddot{\gamma}(t)$$

donde D indica la diferencial de una aplicación de H en \mathbb{R} . Evaluando esta expresión en $t = 0$:

$$0 < \frac{d^2}{dt^2} l^2(\gamma_t)|_{t=0} = \dot{\gamma}(0)^t D^2 \left(d^2(f, \bar{\gamma}(0)) \right) \dot{\gamma}(0) \quad (2.7)$$

Esto es, la forma cuadrática:

$$\begin{aligned} F &\longrightarrow \mathbb{R} \\ v &\longmapsto v^t D^2 \left(d^2(f, \Upsilon(f)) \right) v \end{aligned}$$

es definida positiva, pues la desigualdad (2.7) no depende de la geodésica $\bar{\gamma}$ elegida. De esta manera, $Hess^{d^2(f, \cdot), \Upsilon(f)}$ es definido positivo tal y como queríamos ver. □

Queda así visto que, siempre que nos acerquemos convenientemente a una subvariedad de una variedad Riemanniana de dimensión finita o bien del espacio de Hilbert de las funciones de cuadrado integrable, podemos encontrar

pares de puntos (p_0, θ_0) verificando las hipótesis del teorema 2.2.1. En estos dos ejemplos destacados hemos construido de forma explícita la proyección sobre una subvariedad de un punto convenientemente próximo a ésta. El teorema 2.2.1 prueba que existe un entorno, U , de cualquiera de estos puntos donde la proyección está bien definida. A priori sólo podemos afirmar que $2R^* < \text{diam}(U)$ pero esperamos, en futuras investigaciones esclarecer la relación entre estas dos cantidades y saber dado un punto ϑ de la subvariedad cuál es el radio límite, R^{lim} para el que los puntos de $B(\vartheta, R^{lim})$ admitan, y de manera única, proyección sobre la subvariedad.

Capítulo 3

Familia invariante en orden

Un concepto fundamental en cuestiones de invariancia es el de familia de distribuciones de probabilidad invariante frente a la acción de un grupo. A partir de este concepto se construyen otros como el de estimador equivariante o test invariante. Sin embargo no todas las familias resultan ser invariantes frente a la acción de algún grupo. En este capítulo nos proponemos estudiar cuantitativamente el grado de no invariancia de una familia de distribuciones de probabilidad respecto a una familia arbitraria de transformaciones.

3.1 Perturbaciones en el espacio muestral

Sea $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ un modelo estadístico paramétrico, esto es (Ω, \mathcal{A}) es un espacio medible y $\{P_\theta \mid \theta \in \Theta\}$ es una familia no vacía de distribuciones de probabilidad en (Ω, \mathcal{A}) . Supondremos que Θ , el espacio de parámetros, es una variedad \mathcal{C}^∞ y escribiremos \mathcal{P}_Θ en lugar de $\{P_\theta \mid \theta \in \Theta\}$.

Sea $\mathcal{G}_T = \{g_t \mid t \in T\}$, donde T es un intervalo de \mathbb{R} , una familia de transformaciones de Ω en Ω , biyectivas, bimedibles y conteniendo a la identidad. Notemos como $e \in T$ aquel elemento para el que g_e sea la identidad sobre Ω . Necesitaremos además que Ω sea un espacio topológico pues vamos a exigir a \mathcal{G}_T que $\lim_{t \rightarrow e} g_t w = w$ P_θ - c.s., $\forall \theta \in \Theta$.

Observación 3.1.1. La elección de un intervalo T de \mathbb{R} para etiquetar la familia \mathcal{G}_T se puede extender tomando un subconjunto conexo de cualquier espacio topológico. Se ha optado por \mathbb{R} porque esta elección hace más simple la escritura de las definiciones y resultados que se van a presentar y no suponen pérdida de generalidad.

Sea ahora X una variable aleatoria con valores en Ω . Cada $g_t \in \mathcal{G}$ es una transformación medible de Ω en Ω por lo que $g_t X$ es también una variable aleatoria que toma el valor $g_t x$ cuando $X = x$. Cabe preguntarse por su

distribución de probabilidad. Con las notaciones habituales de teoría de la medida, si $X \sim P_\theta$ escribiremos $P_\theta^{g_t}$ para la ley de la variable aleatoria $g_t X$. En el caso de que \mathcal{G}_T sea un grupo y la familia de distribuciones de probabilidad $\mathcal{P}_\Theta := \{P_\theta \mid \theta \in \Theta\}$ invariante frente a la acción de \mathcal{G}_T , existiría $\theta' \in \Theta$ tal que $g_t X$ seguiría la distribución $P_{\theta'}$. Sin embargo, esta situación no tiene porque ser la habitual. En este trabajo analizaremos la posibilidad de que la variable $g_t X$ siga una ley no enmarcada en el modelo de partida pero tampoco permitiremos que se aleje *demasiado* de él. El significado del término *demasiado* quedará concretado más adelante.

Introducimos ahora el marco necesario para modelizar nuestro problema. Sea $(\Omega, \mathcal{A}, \{P_\nu \mid \nu \in \Delta\})$ un modelo estadístico paramétrico identificable, esto es $\forall \nu, \nu' \in \Delta$ de manera que $\nu \neq \nu'$, también se cumpla que $P_\nu \neq P_{\nu'}$. Supondremos, en nuestro contexto, que:

$$\forall t \in T, \theta \in \Theta \text{ existe un único } \nu \in \Delta \text{ de manera que } P_\theta^{g_t} = P_\nu \quad (3.1)$$

Fijado $t \in T$, la condición (3.1) nos lleva a introducir la siguiente familia de aplicaciones:

$$\begin{aligned} K_t: \Theta &\longrightarrow \Delta \\ \theta &\longmapsto \nu \end{aligned}$$

donde ν viene dado por (3.1). Impondremos la condición adicional de que la aplicación:

$$\begin{aligned} K_e: \Theta &\longrightarrow \Delta \\ \theta &\longmapsto \theta \end{aligned}$$

sea una inmersión global. De esta manera, Θ es una subvariedad de Δ . Recordemos que la definición de subvariedad puede encontrarse en el apéndice correspondiente a Geometría Diferencial de este trabajo.

Notación 3.1.1. Nos referiremos a Δ como la *variedad ambiente* o *variedad ampliada*.

El problema de la determinación del par (Δ, δ) , en una situación general, es un tema abierto que se pretende analizar en posteriores trabajos. En muchas ocasiones, aunque no necesariamente siempre, un candidato natural a desempeñar este papel parece ser el conjunto $T \times \Theta$ junto con el cuadrado de la métrica informativa en aquellas situaciones en las que pueda asegurarse que ésta existe. El siguiente ejemplo sobre modelos paramétricos finito-dimensionales está dedicado a este aspecto.

Ejemplo 3.1.1. Sea (Ω, \mathcal{A}) un espacio medible y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia de distribuciones de probabilidad en (Ω, \mathcal{A}) donde Θ , el espacio de parámetros, es una n -variedad conexa de clase \mathcal{C}^∞ . Finalmente, sea $\mathcal{G}_T = \{g_t \mid t \in T\}$, una familia de transformaciones de Ω en Ω cumpliendo las hipótesis de la sección.

Supongamos que exista una medida σ -finita de referencia, μ , en (Ω, \mathcal{A}) de manera que $\forall (t, \theta) \in T \times \Theta$, $\mu \gg P_\theta^{g_t}$ y sea f una función:

$$\begin{aligned} f : \Omega \times (T \times \Theta) &\longmapsto \mathbb{R} \\ (w, t, \theta) &\longmapsto f(w; t, \theta) \end{aligned}$$

de manera que:

1. Fijado $w \in \Omega$, la aplicación:

$$\begin{aligned} f(w; \cdot, \cdot) : T \times \Theta &\longrightarrow \mathbb{R} \\ (t, \theta) &\longmapsto f(w; t, \theta) \end{aligned}$$

sea de clase \mathcal{C}^∞ .

2. $\forall (t, \theta) \in T \times \Theta$, $f(\cdot; t, \theta)$ sea una versión de la derivada de Radon–Nikodym de P_θ respecto a μ .

Identifiquemos, con un cierto abuso de notación, los puntos de la variedad Θ con sus coordenadas y denotemos $t = \theta^0$. Observemos que $T \times \Theta$ es una variedad Hausdorff y conexa, pues tanto T como Θ lo son y estas propiedades se conservan al realizar un producto cartesiano. Supongamos además las siguientes condiciones adicionales:

1. Para cualquier carta local en (t, θ) las funciones:

$$\frac{\partial}{\partial \theta^0} \log f(\cdot; (t, \theta)), \frac{\partial}{\partial \theta^1} \log f(\cdot; (t, \theta)), \dots, \frac{\partial}{\partial \theta^n} \log f(\cdot; (t, \theta))$$

son linealmente independientes y pertenecen a $L^2(f(\cdot; (t, \theta))d\mu)$.

2. Las derivadas:

$$\frac{d}{dt}, \frac{\partial}{\partial \theta^i}, \frac{\partial^2}{\partial \theta^i \partial \theta^j}, \frac{\partial^3}{\partial \theta^i \partial \theta^j \partial \theta^k}, \dots, \quad i, j, k = 0, \dots, n$$

y la integración respecto a $d\mu$ de $f(\cdot; (t, \theta))$ se puedan intercambiar cuando sea necesario.

Bajo estas condiciones, podemos afirmar que la variedad $T \times \Theta$ tiene una estructura Riemanniana dada por su métrica informativa, tal y como puede consultarse en Amari [2], Atkinson y Mitchell [5], Burbea [9] o bien Burbea y Rao [11] entre muchos otros. Sea ρ la distancia Riemanniana en esta variedad. Tomando $\Delta = T \times \Theta$ y $\delta = \rho^2$ queda probada, bajo ciertas condiciones de regularidad, la existencia del par (Δ, δ) en el caso de modelos paramétricos finito-dimensionales.

Para el caso general de modelos paramétricos dominados, también se podría considerar como variedad Δ el espacio de Hilbert, H , de las funciones de cuadrado integrable y como discrepancia la distancia de Hellinger al cuadrado pues cualquier densidad de probabilidad puede ser identificada con dos veces su raíz cuadrada positiva, que es un elemento de H . No analizaremos ahora el problema de la elección de la variedad ampliada y de la discrepancia y esperaremos a la sección 3.3 donde va a ser estudiado con más detalle.

Continuando con las definiciones y notaciones necesarias para modelizar el problema de la proyección, denotemos $K_t(\theta)$ como $g_t^*\theta$. Podemos introducir ahora:

Definición 3.1.1. Sea $\theta \in \Theta$. Diremos que $P_{g_t^*\theta}$ es la *perturbación de P_θ por $g_t \in \mathcal{G}_T$* .

Definición 3.1.2. Diremos que $\{P_{g_t^*\theta} \mid \theta \in \Theta\}$ es la *perturbación de la familia $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ por $g_t \in \mathcal{G}_T$* .

Identifiquemos las distribuciones de probabilidad con sus etiquetas en el espacio de parámetros y en la variedad ampliada respectivamente. Recordemos que Θ está incluida en Δ al ser una subvariedad de Δ por hipótesis. Por lo tanto tiene sentido preguntarse por $\delta(g_t^*\theta, \Theta) = \inf_{\theta' \in \Theta} \delta(g_t^*\theta, \theta')$ o, más aún, si existe $\theta' \in \Theta$ de manera que $\delta(g_t^*\theta, \Theta) = \delta(g_t^*\theta, \theta')$. En realidad, lo que queremos saber si existe un elemento de la familia original \mathcal{P}_Θ que esté a discrepancia mínima de la perturbación.

Los resultados 2.3.1, 2.3.2 y 2.3.4, 2.3.5, obtenidos en el capítulo 2, prueban que para variedades Riemannianas de dimensión finita, tomando como discrepancia la distancia de Rao al cuadrado, y para el espacio de Hilbert H de las funciones de cuadrado integrable y discrepancia la distancia de Hellinger al cuadrado, si las perturbaciones son cercanas a la identidad, existe $\theta' \in \Theta$ tal que:

$$\delta(g_t^*\theta, \Theta) = \delta(g_t^*\theta, \theta')$$

donde δ sería la discrepancia elegida en cada caso.

Notación 3.1.2. Al elemento θ' de la expresión anterior lo denotaremos por $\overline{g}_t\theta$ y fijado $t \in T$, la aplicación $\Theta \ni \theta \mapsto \overline{g}_t\theta$ será la *transformación inducida por $g_t \in \mathcal{G}_T$* .

Observemos que la proyección es la medida probabilística dentro del modelo original que más se aproxima, según la discrepancia, a la medida transformada.

En una situación general, como sería la analizada en el resultado 2.2.2, debemos partir de un par (p_0, θ_0) verificando las hipótesis del teorema, esto es θ_0 es el único elemento de Θ para el que $\delta(p_0, \Theta) = \delta(p_0, \theta_0)$ y $Hess^{\delta(p_0, \cdot), \theta_0}$ es definido positivo. La primera hipótesis se tendría si tomáramos como punto (p_0, θ_0) un par de la forma (θ_0, θ_0) donde $\theta_0 \in \Theta$. Respecto a la condición sobre el Hessiano necesitaríamos conocer propiedades adicionales sobre la función $\delta(p_0, \cdot)|_{\Theta}$.

De esta manera, nos restringiremos para todo lo que sigue a variedades Riemannianas de dimensión finita o al espacio de Hilbert H e indicaremos como Δ la variedad ampliada en cualquiera de estos dos casos. δ será entonces la discrepancia elegida en cada una de estas dos situaciones: la distancia de Rao al cuadrado para variedades Riemannianas de dimensión finita y la de Hellinger, también al cuadrado, para H .

Observemos que la curva $t \mapsto g_t^*\theta$ está definida para todo $t \in T$ (aquí se entiende que θ es arbitrario pero fijo) mientras que la curva $t \mapsto \overline{g}_t\theta$ únicamente está definida para t próximos a e pero dependiendo de θ . Más detalladamente, $\exists R_\theta > 0$ tal que si $|t - e| < R_\theta$ el elemento $\overline{g}_t\theta$ está bien definido. Finalmente, observemos que $\lim_{t \rightarrow e} \delta(g_t^*\theta, \overline{g}_t\theta) = 0$ pues se ha exigido a \mathcal{G}_T que $\lim_{t \rightarrow e} g_t w = w$ $P_\theta - c.s.$, $\forall \theta \in \Theta$. Esto implica que $\lim_{t \rightarrow e} g_t^*\theta = \lim_{t \rightarrow e} \overline{g}_t\theta = \theta$, $\forall \theta \in \Theta$. De aquí, recordando que δ es de clase \mathcal{C}^∞ (de hecho únicamente necesitamos continuidad), se tiene que $\lim_{t \rightarrow e} \delta(g_t^*\theta, \overline{g}_t\theta) = 0$.

Estas consideraciones previas nos llevan a plantear un nuevo problema abierto a estudio. Podemos preguntarnos cuándo existe un intervalo I de manera que la transformación:

$$I \ni t \mapsto \overline{g}_t\theta \quad (3.2)$$

esté bien definida para todo $\theta \in \Theta$. Un caso claro en el que esto ocurre es cuando el espacio de parámetros Θ sea compacto. La situación general no es evidente y se intentará en trabajos posteriores caracterizar el subconjunto de Θ formado por los $\theta \in \Theta$ para los que no existe un intervalo I cumpliendo (3.2).

El teorema 2.2.2 informa que la transformación $\theta \mapsto \overline{g}_t\theta$ es de clase \mathcal{C}^∞ al serlo δ . Sin embargo, observemos que en general no es inversible. La siguiente

figura resume el problema de la invertibilidad.

$$\overline{g_t}\theta \neq g_t^*\theta$$

Figura 3.1: Representación gráfica de la no invertibilidad, en general, de las transformaciones $\theta \mapsto \overline{g_t}\theta$.

El último punto a preguntarse es como difieren $g_t^*\theta$ de $\overline{g_t}\theta$ pues esta diferencia indicará lo alejados que nos encontramos de la situación clásica de invariancia en la que uno y otro elemento coinciden. La siguiente sección está dedicada a este aspecto.

3.2 Familia invariante de orden α_I

Con las notaciones de la sección anterior, un primer intento de cuantificación de la no invariancia de una familia \mathcal{P}_Θ respecto a las transformaciones inducidas en el espacio de parámetros por una familia arbitraria de transformaciones \mathcal{G}_T , es la siguiente.

Definición 3.2.1. Diremos que \mathcal{P}_Θ es *invariante, relativa a la parametrización dada por T , de orden α_R* respecto a la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\alpha_R = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\delta(g_t^*\theta, \overline{g_t}\theta)}{(t - e)^\alpha} = 0, \forall \theta \in \Theta\}$$

Observemos que la definición anterior depende de la parametrización dada por T tal y como se muestra en el siguiente ejemplo.

Ejemplo 3.2.1. Sea X una variable aleatoria siguiendo la distribución de probabilidad $N(\mu, 1)$, $\mu \in \mathbb{R}$ y consideremos el grupo multiplicativo de los reales positivos $\mathcal{G} = \{g_a \mid a > 0\}$ donde $g_ax = ax$, $x \in \mathbb{R}$, $a > 0$.

La variable aleatoria $g_aX = aX$ sigue una ley $N(a\mu, a)$, esto es, la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ no es invariante frente a la acción de \mathcal{G} . Observemos que la perturbación de la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ por un elemento de

\mathcal{G} se incluye, de forma natural, en $\{N(\mu, \sigma) \mid (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+\}$. Podemos considerar entonces como variedad ampliada $\mathbb{R} \times \mathbb{R}_+$ y como discrepancia en este ambiente la distancia Riemanniana, o de Rao, al cuadrado ρ^2 . De esta manera, la perturbación de $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ por $g_a \in \mathcal{G}$ induce una transformación de $\Theta = \{(\mu, 1) \mid \mu \in \mathbb{R}\}$ en $\Delta = \mathbb{R} \times \mathbb{R}_+$ dada por:

$$\begin{aligned} g_a^* : \quad \Theta &\longrightarrow \Delta \\ (\mu, 1) &\longmapsto (a\mu, a) \end{aligned}$$

Fijados $(\mu, 1) \in \Theta$ y un real positivo, a , estamos interesados qué elemento de la familia original está a distancia de Rao mínima de $g_a^*(\mu, 1) = (a\mu, a)$. Esto es, debemos estudiar:

$$\min_{\xi \in \mathbb{R}} \rho^2((a\mu, a), (\xi, 1))$$

La distancia de Rao al cuadrado entre $N(a\mu, a)$ y $N(\xi, 1)$ es (ver Burbea [9]):

$$\rho^2((a\mu, a), (\xi, 1)) = 2 \log^2 \left(\frac{1 + \eta(\xi, \mu, a)}{1 - \eta(\xi, \mu, a)} \right)$$

donde:

$$\eta(\xi, \mu, a) = \left(\frac{(\xi - a\mu)^2 + 2(a - 1)^2}{(\xi - a\mu)^2 + 2(a + 1)^2} \right)^{\frac{1}{2}}$$

y con el convenio de notar $\rho^2((\mu_1, \sigma_1), (\mu_2, \sigma_2))$ como la distancia de Rao, al cuadrado entre $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, $\forall (\mu_1, \sigma_1), (\mu_2, \sigma_2) \in \mathbb{R} \times \mathbb{R}_+$. Observemos que $\rho^2((a\mu, a), (\xi, 1))$ se hace mínima si $\xi = a\mu$, esto es:

$$\min_{\xi \in \mathbb{R}} \rho^2((a\mu, a), (\xi, 1)) = \rho^2((a\mu, a), (a\mu, 1)) = 2 \log^2(a)$$

Así, hemos obtenido una familia de transformaciones en Θ que viene dada por:

$$\begin{aligned} \overline{g}_a : \quad \Theta &\longrightarrow \Theta \\ (\mu, 1) &\longmapsto (a\mu, 1) \end{aligned}$$

Vamos a calcular ahora el orden de invariancia relativo de la familia respecto a \mathcal{G} . Para a cercanas a 1:

$$\rho^2((a\mu, a), (a\mu, 1)) = 2 \log^2(a) = 2 \left(\sum_{j=1}^{+\infty} (-1)^{j+1} \frac{(a-1)^j}{j} \right)^2 =$$

$$= 2 (a - 1)^2 + O((a - 1)^3)$$

De esta manera:

$$\lim_{a \rightarrow 1} \frac{\rho^2((a\mu, a), (a\mu, 1))}{(a - 1)^\alpha} = 0 \text{ si } \alpha < 2$$

Esto es, el orden de invariancia relativo a la parametrización dada por "las a " es 2. Veamos qué ocurre si expresamos $a = (t - 1)^{2\gamma+1} + 1$, con $t \in [0, +\infty)$ y γ un natural positivo. En este caso, debemos calcular:

$$\begin{aligned} \lim_{t \rightarrow 1} \frac{\rho^2\left(\left((t - 1)^{2\gamma+1} + 1\right) \mu, (t - 1)^{2\gamma+1} + 1\right), \left(\left((t - 1)^{2\gamma+1} + 1\right) \mu, 1\right)}{(t - 1)^\alpha} = \\ = \lim_{t \rightarrow 1} \frac{2 \log^2((t - 1)^{2\gamma+1} + 1)}{(t - 1)^\alpha} \end{aligned}$$

Para t suficientemente cercanas a 1:

$$2 \log^2((t - 1)^{2\gamma+1} + 1) = 2 (t - 1)^{4\gamma+2} + O((t - 1)^{6\gamma+3})$$

De esta manera, para la nueva parametrización, el orden de invariancia relativo es $4\gamma + 2$, que depende claramente de la elección a la hora de etiquetar la familia de transformaciones \mathcal{G}_T .

Una posibilidad para eliminar la dependencia anteriormente citada sería considerar como factor infinitesimal de comparación $\delta(g_t^*\theta, \theta)$ y analizar el comportamiento de $\delta(g_t^*\theta, \overline{g_t}\theta)$ respecto a $\delta(g_t^*\theta, \theta)$ cuando $t \rightarrow e$.

Definición 3.2.2. Diremos que \mathcal{P}_Θ es *invariante intrínseca de orden α_I* respecto a la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\alpha_I = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\delta(g_t^*\theta, \overline{g_t}\theta)}{\delta(g_t^*\theta, \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

En la definición anterior, es posible que el supremo sea en realidad un máximo. Esto nos lleva a introducir:

Definición 3.2.3. Diremos que \mathcal{P}_Θ es *invariante intrínseca de orden máximo α_I* respecto a la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\alpha_I = \max\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\delta(g_t^*\theta, \overline{g_t}\theta)}{\delta(g_t^*\theta, \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

Observemos que si una familia es invariante intrínseca de orden máximo α_I también es invariante intrínseca de orden α_I . El recíproco en general no tiene porque ser cierto.

En primer lugar se debe verificar que efectivamente el orden de invariancia intrínseco no cambia frente a reparametrizaciones del parámetro t . De esta manera, el orden de invariancia únicamente parece depender de la medida de discrepancia escogida. Recordemos que este problema se analizará con detalle en la sección 3.3 estableciéndose allí que para una amplia clase de discrepancias el orden de invariancia únicamente va a depender de la familia de distribuciones de probabilidad en estudio y obviamente de la familia de transformaciones en el espacio de parámetros.

Pasamos a demostrar que el orden de invariancia de una familia no se ve alterado frente a reparametrizaciones de $t \in T$.

Proposición 3.2.1. *Sea $S \subset \mathbb{R}$ y $\phi : S \longrightarrow T$ una aplicación biyectiva y bicontinua en $e' \in S$ siendo $\phi(e') = e$.*

Entonces:

$$\alpha_I = \sup\{\alpha \geq 0 \mid \lim_{s \rightarrow e'} \frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)} \theta})}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

Demostración. Introducimos:

$$\alpha_{\phi I} = \sup \{ \alpha \geq 0 \mid \lim_{s \rightarrow e'} \frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)} \theta})}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} = 0, \forall \theta \in \Theta \}$$

Queremos ver que $\alpha_{\phi I} = \alpha_I$. Probaremos la doble desigualdad. Sea $\alpha \geq 0$ tal que:

$$\lim_{t \rightarrow e} \frac{\delta(g_t^* \theta, \overline{g_t \theta})}{\delta(g_t^* \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta$$

Veremos que α también cumple:

$$\lim_{s \rightarrow e'} \frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)} \theta})}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta \quad (3.3)$$

condición que implica, utilizando la definición de supremo de un conjunto, que $\alpha_{\phi I} \geq \alpha_I$.

Veamos que la condición (3.3) se cumple. Sea $\theta \in \Theta$. Por definición de límite, dado $\forall \epsilon_1 > 0 \exists \eta_1 > 0$ tal que:

$$\frac{\delta(g_t^* \theta, \overline{g_t \theta})}{\delta(g_t^* \theta, \theta)^\alpha} < \epsilon_1, \text{ siempre que } |t - e| < \eta_1 \quad (3.4)$$

Por otra parte, la función $s \mapsto \phi(s)$ es continua en $s = e'$, donde recordemos que $\phi(e') = e$. De esta manera, dado $\epsilon_2 > 0$ existirá $\eta_2 > 0$ tal que:

$$|\phi(s) - \phi(e')| < \epsilon_2, \text{ siempre que } |s - e'| < \eta_2 \quad (3.5)$$

Sea ahora $\epsilon > 0$ y hagamos $t = \phi(s)$, $e = \phi(e')$ en (3.4). Existe $\eta_1 > 0$ tal que $|\phi(s) - \phi(e')| < \eta_1$ implica que:

$$\frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)}} \theta)}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} < \epsilon$$

Para el $\eta_1 > 0$ anterior, existe $\eta_2 > 0$ dado por (3.5) de manera que:

$$|s - e'| < \eta_2, \text{ implica que } |\phi(s) - \phi(e')| < \eta_1$$

Si llamamos $\eta = \eta_2$ y recopilamos toda la información, hemos demostrado que $\forall \epsilon > 0 \exists \eta > 0$ tal que $|s - e| < \eta$ implica que:

$$\frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)}} \theta)}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} < \epsilon$$

esto es:

$$\lim_{t \rightarrow e} \frac{\delta(g_{\phi(s)}^* \theta, \overline{g_{\phi(s)}} \theta)}{\delta(g_{\phi(s)}^* \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta$$

pues θ era un elemento arbitrario pero fijo de Θ .

Para demostrar la desigualdad contraria, basta que reproduzcamos este razonamiento anterior pero con ϕ^{-1} en el papel de ϕ .

□

Teniendo en cuenta la proposición 3.2.1, cuantificaremos el orden de invariancia de una familia de distribuciones de probabilidad por medio del orden intrínseco ya que de esta manera, eliminaremos la dependencia respecto a la parametrización que elijamos para etiquetar las transformaciones en el espacio muestral.

3.3 Variedad ambiente

Una vez verificada la condición intrínseca de α_I , cabe preguntarse si la elección de la variedad ampliada o ambiente influye en los órdenes de invariancia. En general no puede afirmarse nada al respecto, pues los órdenes de invariancia dependen de la medida de discrepancia escogida. Sin embargo, en esta

sección veremos que existe una amplia clase de discrepancias para las que así es.

Vamos a suponer que Δ , la variedad ampliada, es Riemanniana y que la distancia Riemanniana, ρ , es la dada por la matriz de información de Fisher. Requeriremos, además que \mathcal{P}_Θ sea una familia paramétrica de distribuciones de probabilidad absolutamente continuas respecto μ , una medida σ -finita en (Ω, \mathcal{A}) , y que Θ , el espacio de parámetros es una subvariedad Riemanniana de Δ de dimensión finita.

Consideraremos como medida de discrepancia natural la propia distancia Riemanniana al cuadrado y supondremos que \mathcal{P}_Θ es invariante de orden α_I respecto a una familia de transformaciones \mathcal{G}_T . Esta elección de discrepancia recoge las propiedades geométricas del modelo y veremos que cumple que el orden de invariancia de una familia paramétrica de probabilidades respecto a ρ es el mismo que respecto a una amplia clase de divergencias que pasamos a introducir.

Sean P_{α_1} y P_{α_2} medidas de probabilidad en (Ω, \mathcal{A}) absolutamente continuas respecto μ y sea $p(\cdot; \alpha_i) = \frac{dP_{\alpha_i}}{d\mu}$, $i = 1, 2$. Consideremos la clase de divergencias de Csiszár [14], cuya expresión viene dada por:

$$D(P_{\alpha_1}, P_{\alpha_2}) = \int p(x; \alpha_1) f\left(\frac{p(x; \alpha_2)}{p(x; \alpha_1)}\right) \mu(dx)$$

siendo f es una función estrictamente convexa y de clase \mathcal{C}^2 . Estas divergencias son conocidas también como f -divergencias.

El siguiente resultado va a resultar un elemento fundamental en esta sección al permitir comparar las distancias Riemanniana y Euclídea en el límite y puede encontrarse demostrado en el capítulo 1, sección 9, de Helgason [22]. En el lema 3.3.1 se hace uso del concepto de entorno normal de un punto que puede encontrarse también en Helgason [22] o bien en el apéndice correspondiente a Geometría Diferencial, definición A.7.7.

Lema 3.3.1. *Sea (Δ, ρ) una variedad Riemanniana y $B_r(p)$ un entorno normal dado por el lema A.7.3. Sean $a, b \in B_r(p)$ y A, B los únicos puntos en $T_p(\Delta)$ cumpliendo las relaciones:*

$$\exp_p(A) = a, \quad \exp_p(B) = b, \quad \|A\| \leq r(p), \quad \|B\| \leq r(p).$$

Entonces:

$$\lim_{(a,b) \rightarrow (p,p)} \frac{\|A - B\|}{\rho(a, b)} = 1$$

Con todas estas ideas presentes, demostraremos el resultado principal de esta sección.

Proposición 3.3.2. *Con las hipótesis de la sección, el orden de invariancia de \mathcal{P}_Θ no depende de la medida de discrepancia escogida en la variedad ambiente.*

Demostración. Con las notaciones del capítulo, sea $\mathcal{M}(\Omega, \mathcal{A}, \mu)$ el conjunto de todas las medidas de probabilidad en (Ω, \mathcal{A}) , absolutamente continuas respecto a una medida σ -finita μ en (Ω, \mathcal{A}) . Sea D_f una f -divergencia entre elementos de $\mathcal{M}(\Omega, \mathcal{A}, \mu)$. Introducimos:

$$\alpha_{f-I} = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{D_f(g_t^* \theta, \bar{g}_t \theta)}{D_f(g_t^* \theta, \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

Queremos demostrar que $\alpha_{f-I} = \alpha_I$. Necesitaremos realizar una serie de cálculos previos. Sean θ_1 y $\theta_2 \in \Delta$ y $\Delta\theta = \theta_2 - \theta_1$. Con abuso de notación, omitiremos la presencia de una n -carta local. Siguiendo las ideas desarrolladas en Burbea y Rao [11] (que también pueden encontrarse en Amari et al. [3]) podemos expresar:

$$D_f(\theta_1, \theta_2) = \frac{1}{2} f''(1) \sum_{\alpha, \beta=1}^n g_{\alpha\beta}(\theta_1) \Delta\theta_\alpha \Delta\theta_\beta + O(\|\Delta\theta\|^3) \quad (3.6)$$

siendo $g_{\alpha\beta}(\theta_1)$ los elementos la matriz de información de Fisher en θ_1 y la derivada segunda de f evaluada en 1 diferente de 0 pues D_f era una f -divergencia.

Tomemos ahora $\theta_1 = g_t^* \theta$ y $\theta_2 = \bar{g}_t \theta$ y sea $\Delta\theta(t) = \bar{g}_t \theta - g_t^* \theta$. Con ayuda de la expresión (3.6) podemos escribir:

$$\lim_{t \rightarrow e} \frac{D_f(g_t^* \theta, \bar{g}_t \theta)}{\rho^2(g_t^* \theta, \bar{g}_t \theta)} = \frac{1}{2} f''(1) + \lim_{t \rightarrow e} \frac{\|\Delta\theta(t)\|^3}{\rho^2(g_t^* \theta, \bar{g}_t \theta)}$$

En un espacio normado de dimensión finita todas las normas son equivalentes, por lo tanto existe una constante positiva a tal que:

$$\lim_{t \rightarrow e} \frac{\|\Delta\theta(t)\|}{\rho(g_t^* \theta, \bar{g}_t \theta)} \leq a \lim_{t \rightarrow e} \frac{\|\exp_\theta^{-1} \bar{g}_t \theta - \exp_\theta^{-1} g_t^* \theta\|}{\rho(g_t^* \theta, \bar{g}_t \theta)}$$

Por otra parte, el lema 3.3.1 informa que:

$$\lim_{t \rightarrow e} \frac{\|\exp_\theta^{-1} \bar{g}_t \theta - \exp_\theta^{-1} g_t^* \theta\|}{\rho(g_t^* \theta, \bar{g}_t \theta)} = 0$$

Así ha de ser:

$$0 \leq \lim_{t \rightarrow e} \frac{\|\Delta\theta(t)\|^3}{\rho^2(g_t^*\theta, \bar{g}_t\theta)} \leq a \lim_{t \rightarrow e} \frac{\|\exp_{\theta}^{-1} \bar{g}_t\theta - \exp_{\theta}^{-1} g_t^*\theta\|}{\rho^3(g_t^*\theta, \bar{g}_t\theta)} \rho(g_t^*\theta, \bar{g}_t\theta) = 0$$

y por lo tanto se cumplirá que:

$$\lim_{t \rightarrow e} \frac{D_f(g_t^*\theta, \bar{g}_t\theta)}{\rho^2(g_t^*\theta, \bar{g}_t\theta)} = \frac{1}{2} f''(1) \neq 0 \quad (3.7)$$

Veamos ahora que $\alpha_{f-I} = \alpha_I$. Sea $\alpha < \alpha_{f-I}$ y $\theta \in \Theta$. Entonces:

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^*\theta, \bar{g}_t\theta)}{\rho^2(g_t^*\theta, \theta)^\alpha} = \lim_{t \rightarrow e} \frac{\rho_A^2(g_t^*\theta, \bar{g}_t\theta)}{D_f(g_t^*\theta, \bar{g}_t\theta)} \cdot \frac{D_f(g_t^*\theta, \bar{g}_t\theta)}{D_f(g_t^*\theta, \theta)^\alpha} \cdot \frac{D_f(g_t^*\theta, \theta)^\alpha}{\rho_A^2(g_t^*\theta, \theta)^\alpha} = 0$$

según (3.7) y recordando la definición de α_{f-I} . Al ser $\theta \in \Theta$ arbitrario y como consecuencia de la definición de supremo de un conjunto ha de ser $\alpha_{f-I} \leq \alpha_I$.

Probemos la desigualdad contraria. De nuevo, sea $\theta \in \Theta$ y consideremos $\alpha < \alpha_I$. Entonces:

$$\lim_{t \rightarrow e} \frac{D_f(g_t^*\theta, \bar{g}_t\theta)}{D_f(g_t^*\theta, \theta)^\alpha} = \lim_{t \rightarrow e} \frac{D_f(g_t^*\theta, \bar{g}_t\theta)}{\rho^2(g_t^*\theta, \bar{g}_t\theta)} \cdot \frac{\rho^2(g_t^*\theta, \bar{g}_t\theta)}{\rho^2(g_t^*\theta, \theta)^\alpha} \cdot \frac{\rho^2(g_t^*\theta, \theta)^\alpha}{D_f(g_t^*\theta, \theta)^\alpha} = 0$$

utilizando nuevamente la expresión (3.7) y la definición de α_I . Como $\theta \in \Theta$ es arbitrario, $\alpha_I \leq \alpha_{f-I}$. □

Queda así demostrado que podríamos haber escogido como medida de discrepancia cualquier f -divergencia, en particular la distancia de Hellinger, sin que los órdenes de invariancia de la familia se vieran afectados por esta elección. Observemos que este resultado es razonable pues la distancia de Hellinger es, salvo un factor de proporcionalidad que no afecta al cálculo de los órdenes, la de Rao en un contexto más amplio.

La elección de una clase razonable de medidas de discrepancia es un problema abierto. En este trabajo se ha optado por seleccionar f -divergencias pues únicamente dependen de las medidas de probabilidad que estemos analizando y no del contexto en el que se hayan incluido. Ésta es una posición metodológica que puede ser conveniente cuando la única información de la que se dispone es la de existencia de un cierto modelo paramétrico. Ante información adicional es posible que debiéramos amplificar el modelo, o bien construir una discrepancia *ad hoc*. Tal y como se ha comentado anteriormente, es un tema que se pretende analizar en futuros trabajos.

3.4 Ejemplos.

Esta sección está dedicada a calcular los órdenes de invariancia intrínsecos para algunos ejemplos concretos.

Ejemplo 3.4.1. Un caso interesante a considerar es el de una familia paramétrica de distribuciones de probabilidad invariante frente a la acción de un grupo. Observemos que en esta situación, $\forall t \in T, \theta \in \Theta$ se cumple que $g_t^* \theta = \overline{g}_t \theta$. Entonces, para cualquier medida de discrepancia δ y $\forall \alpha \geq 0$:

$$\lim_{t \rightarrow e} \frac{\delta(g_t^* \theta, \overline{g}_t \theta)}{\delta(g_t^* \theta, \theta)^\alpha} = 0, \forall \theta \in \Theta$$

Esto es, para la situación clásica de invariancia $\alpha_I = +\infty$.

Ejemplo 3.4.2. Volviendo al ejemplo 3.2.1 calculemos ahora el orden de invariancia intrínseco de la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ respecto al grupo multiplicativo de los reales, considerando como discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$ la distancia de Rao al cuadrado, ρ^2 .

Ya habíamos visto que:

$$\rho^2((a\mu, a), (a\mu, 1)) = 2 \log^2(a) = 2(a-1)^2 + O((a-1)^3)$$

Por otra parte:

$$\rho^2((a\mu, a), (\mu, 1)) = 2 \left(\log^2 \left(\frac{1 + \eta(\mu, \mu, a)}{1 - \eta(\mu, \mu, a)} \right) \right)^2$$

donde recordemos que:

$$\eta(\mu, \mu, a) = \left(\frac{(\mu - a\mu)^2 + 2(a-1)^2}{(\mu - a\mu)^2 + 2(a+1)^2} \right)^{\frac{1}{2}}$$

De aquí se obtiene que:

$$\rho^2((a\mu, a), (\mu, 1)) = (\mu^2 + 2)(a-1)^2 + O((a-1)^3)$$

Con lo que podemos afirmar:

$$\lim_{a \rightarrow 1} \frac{\rho^2((a\mu, a), (a\mu, 1))}{\rho^2((a\mu, a), (\mu, 1))^\alpha} = 0 \text{ si } \alpha < 1$$

esto es, el orden de invariancia intrínseco de la familia respecto al grupo multiplicativo de los reales positivos es 1.

Ejemplo 3.4.3. Observemos que en el ejemplo anterior, podríamos haber tomado como variedad ambiente el espacio de Hilbert de las funciones de cuadrado integrable y como medida de discrepancia la distancia de Hellinger al cuadrado, d_H^2 . En virtud de la proposición 3.3.2 nada debería cambiar respecto a los resultados que se han obtenido. Vamos a verificarlo. Se cumple que:

$$d_H^2((a\mu, a), (\xi, 1)) = 2 - \frac{2}{\sqrt{a}} \sqrt{\frac{2a^2}{a^2 + 1}} \exp^{-\frac{1}{4} \frac{(\xi - a\mu)^2}{a^2 + 1}}$$

donde, con los mismos criterios que para la distancia Riemanniana al cuadrado, $d_H^2((\mu_1, \sigma_1), (\mu_2, \sigma_2))$ indica $d_H^2(N(\mu_1, \sigma_1), N(\mu_2, \sigma_2))$ para elementos del espacio de parámetros $(\mu_1, \sigma_1), (\mu_2, \sigma_2) \in \mathbb{R} \times \mathbb{R}_+$.

Observemos que:

$$\min_{\xi \in \mathbb{R}} d_H^2((a\mu, a), (\xi, 1)) = d_H^2((a\mu, a), (a\mu, 1))$$

Entonces, para a suficientemente cercanas a 1:

$$d_H^2((a\mu, a), (a\mu, 1)) = 2 \left(1 - \sqrt{\frac{2a}{a^2 + 1}}\right) = \frac{1}{2!} (a - 1)^2 + O((a - 1)^3)$$

y:

$$\begin{aligned} d_H^2((a\mu, a), (\mu, 1)) &= 2 - \frac{2}{\sqrt{a}} \sqrt{\frac{2a^2}{a^2 + 1}} \exp^{-\frac{1}{4} \frac{(\xi - a\mu)^2}{a^2 + 1}} = \\ &= \frac{1}{2!} \left(\frac{\mu^2}{2} + 1\right) (a - 1)^2 + O((a - 1)^3) \end{aligned}$$

Cabe destacar que estos resultados concuerdan con que localmente $4 d_H^2 = \rho^2$ pues recordemos que:

$$\rho^2((a\mu, a), (a\mu, 1)) = 2 \log^2(a) = 2 (a - 1)^2 + O((a - 1)^3)$$

y:

$$\rho^2((a\mu, a), (\mu, 1)) = (\mu^2 + 2)(a - 1)^2 + O((a - 1)^3)$$

Finalmente, podemos afirmar que:

$$\lim_{a \rightarrow 1} \frac{d_H^2((a\mu, a), (a\mu, 1))}{d_H^2((a\mu, a), (\mu, 1))^\alpha} = 0 \text{ si } \alpha < 1$$

como cabía esperar.

Ejemplo 3.4.4. Consideremos la familia paramétrica de distribuciones de probabilidad $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$ y sea $\mathcal{G} = \{g_a \mid a \in \mathbb{R}\}$ el grupo aditivo real dado por:

$$\begin{aligned} g_a : \mathbb{R} &\rightarrow \mathbb{R} & a \in \mathbb{R} \\ x &\mapsto a + x \end{aligned}$$

Si $\sigma > 0$ y X es una variable aleatoria siguiendo la ley $N(0, \sigma)$, la variable transformada por $g_a \in \mathcal{G}$ sigue una distribución $N(a, \sigma)$, que no se enmarca en la familia de origen.

La perturbación de la familia $\{N(0, \sigma) \mid \sigma \in \mathbb{R}\}$ está incluida de forma natural en $\{N(\mu, \sigma) \mid (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+\}$. Consideraremos, entonces, como variedad ampliada $\mathbb{R} \times \mathbb{R}_+$ y como medida de discrepancia en este ambiente la distancia de Rao, ρ . Debemos por tanto estudiar:

$$\min_{\xi^2 > 0} \rho^2((a, \sigma), (0, \xi))$$

Volviendo a consultar Burbea [9], se verifica que la distancia de Rao entre las distribuciones $N(a, \sigma)$ y $N(0, \xi)$ se hace mínima si $\xi^2 = \sigma^2 + \frac{a^2}{2}$ y en este caso:

$$\rho^2((a, \sigma), (0, \sqrt{\sigma^2 + \frac{a^2}{2}})) = 2 \log^2\left(\frac{1 + \delta_{12}}{1 - \delta_{12}}\right) = 8 \tanh^{-1}(\delta_{12})$$

siendo:

$$\delta_{12} = \left(\frac{a^2 + 2(\sigma - \sqrt{\sigma^2 + \frac{a^2}{2}})^2}{a^2 + 2(\sigma + \sqrt{\sigma^2 + \frac{a^2}{2}})^2} \right)^{\frac{1}{2}}$$

Si desarrollamos en serie de Taylor en un entorno de 0:

$$\rho^2((a, \sigma), (0, \sqrt{\sigma^2 + \frac{a^2}{2}})) = 4 \frac{a^2}{\sigma^2} + O(a^4) \quad (3.8)$$

Por otra parte, y de nuevo de Burbea [9] se tiene:

$$\rho^2((a, \sigma), (0, \sigma)) = \frac{a^2}{\sigma^2}$$

Si recopilamos toda la información:

$$\lim_{a \rightarrow 0} \frac{\rho^2((a, \sigma), (0, \sqrt{\sigma^2 + \frac{a^2}{2}}))}{\rho^2((a, \sigma), (0, \sigma))^\alpha} = 0 \text{ si } \alpha < 1$$

esto es, el orden de invariancia intrínseco de la familia respecto al grupo aditivo de los reales es 1.

3.5 Transformación inducida inversa aproximada.

De igual manera a como exigimos en la sección 3.3, vamos a suponer que se cumplen las condiciones A1, A2 y A3, que la variedad ambiente Δ es Riemanniana, siendo la métrica la informativa y que el espacio de parámetros Θ es una subvariedad Riemanniana de Δ de dimensión finita. Bajo estas condiciones, ha quedado establecido en la proposición 3.3.2 que podríamos escoger como discrepancia cualquier f -divergencia pues los órdenes de invariancia de una familia no se ven afectados por esta elección. Tomaremos como discrepancia la distancia Riemanniana en Δ pues, como se probó en la proposición 3.3.2, cualquier f -divergencia es comparable en el límite con ella y recordemos que únicamente estamos interesados en perturbaciones infinitesimales de las distribuciones de probabilidad. Además sus propiedades de invariancia en sentido clásico, de las que veremos una aplicación en la proposición 3.5.1, hacen que sea un candidato razonable a discrepancia.

Es bien sabido que en el caso de disponer de una familia de distribuciones de probabilidad invariante frente a la acción de un grupo, las transformaciones inducidas en el espacio de parámetros son un grupo también. Es de esperar que en nuestro caso no se de una situación tan favorable. Sin embargo, podemos preguntarnos si existe algún elemento de la familia de transformaciones inducidas que haga el papel de *inverso de un cierto orden* de \bar{g}_t , $t \in T$. El siguiente resultado prueba que, en efecto, así es.

Proposición 3.5.1. *Sea $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ un modelo estadístico paramétrico siendo Θ , el espacio de parámetros una variedad \mathcal{C}^∞ . Con las notaciones de la sección 3.1, sea (Δ, ρ) la variedad ampliada, que supondremos Riemanniana, e indiquemos por ρ_Θ la distancia Riemanniana en Θ inducida por la matriz de la métrica de Δ .*

Supongamos que $\{P_\theta \mid \theta \in \Theta\}$ es invariante intrínseca de orden α_I respecto a un grupo de transformaciones bimedibles de Ω en Ω , \mathcal{G}_T e introducimos:

$$\alpha_{inv} = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\rho_\Theta^2(\bar{g}_t \bar{g}_t^{-1} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

Entonces:

$$\alpha_{inv} \geq \alpha_I$$

Demostración. Sea $\theta \in \Theta$ y $t \in T$. Consideremos $\alpha < \alpha_I$. Se tiene:

$$\lim_{t \rightarrow e} \frac{\rho_\Theta^2(\bar{g}_t \bar{g}_t^{-1} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha} = \lim_{t \rightarrow e} \frac{\rho_\Theta^2(\bar{g}_t \bar{g}_t^{-1} \theta, \theta)}{\rho^2(\bar{g}_t \bar{g}_t^{-1} \theta, \theta)} \frac{\rho^2(\bar{g}_t \bar{g}_t^{-1} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha}$$

Se verifica que:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t g_t^{-1}} \theta, \theta)}{\rho^2(\overline{g_t g_t^{-1}} \theta, \theta)} = 1 \quad (3.9)$$

pues si notamos como \exp la aplicación exponencial correspondiente a la conexión de Levi-Civita en Δ y $\overline{\exp}$ la correspondiente a la conexión inducida en Θ por métrica informativa de Δ , se cumple que:

$$\lim_{t \rightarrow e} \frac{\|\exp_{\theta}^{-1} \overline{g_t g_t^{-1}} \theta\|}{\|\overline{\exp}_{\theta}^{-1} \overline{g_t g_t^{-1}} \theta\|} = 1$$

Por otra parte, las distancias Riemanniana y Euclídea siempre son comparables en el límite, tal y como se indicó en el lema 3.3.1, con lo que la expresión (3.9) es cierta.

De esta manera, basta que demostremos:

$$\lim_{t \rightarrow e} \frac{\rho^2(\overline{g_t g_t^{-1}} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^{\alpha}} = 0$$

Utilizando la desigualdad triangular:

$$\rho(\overline{g_t^{-1}} \overline{g_t} \theta, \theta) \leq \rho(\overline{g_t^{-1}} \overline{g_t} \theta, g_t^{-1*} \overline{g_t} \theta) + \rho(g_t^{-1*} \overline{g_t} \theta, \theta)$$

Se cumple que ρ^2 es invariante, en sentido clásico, frente a la acción de un grupo. Es un resultado conocido del cual se ha incluido una demostración detallada en el capítulo 4 este trabajo, proposición 4.2.1, pues este es el capítulo dedicado a analizar la invariancia de la función de pérdida. Aún así, nos vamos a permitir la licencia de utilizar la invariancia de ρ^2 en sentido clásico frente a la acción de un grupo en la siguiente expresión:

$$\rho(\overline{g_t^{-1}} \overline{g_t} \theta, g_t^{-1*} \overline{g_t} \theta) + \rho(g_t^{-1*} \overline{g_t} \theta, \theta) = \rho(\overline{g_t^{-1}} \overline{g_t} \theta, g_t^{-1*} \overline{g_t} \theta) + \rho(\overline{g_t} \theta, g_t^* \theta)$$

Recordando ahora la definición de proyección obtenemos que:

$$\rho(g_t^{-1*} \overline{g_t} \theta, \overline{g_t^{-1}} \overline{g_t} \theta) \leq \rho(g_t^{-1*} \overline{g_t} \theta, \theta) = \rho(g_t^* \theta, \overline{g_t} \theta)$$

Así:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t g_t^{-1}} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^{\alpha}} \leq \lim_{t \rightarrow e} 4 \frac{\rho_{\Theta}^2(g_t^* \theta, \overline{g_t} \theta)}{\rho^2(g_t^* \theta, \theta)^{\alpha}} = 0$$

al ser $\alpha < \alpha_I$ y $\{P_{\theta} \mid \theta \in \Theta\}$ es invariante intrínseca de orden α_I . Por ser $\theta \in \Theta$ arbitrario pero fijo, esta condición prueba que:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t g_t^{-1}} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta$$

esto es, $\alpha_I \leq \alpha_{inv}$.

□

Observemos que nada cambia en la demostración anterior si cambiamos g_t^{-1} por g_t . De esta manera:

$$\alpha_{inv} = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t^{-1}} \overline{g_t} \theta, \theta)}{\rho^2(g_t^{-1*} \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta\}$$

y también:

$$\alpha_{inv} = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t^{-1}} \overline{g_t} \theta, \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha} = 0, \quad \forall \theta \in \Theta\}$$

pues $\rho^2(g_t^* \theta, \theta) = \rho^2(g_t^{-1*} \theta, \theta)$, $\forall t \in T$, $\forall \theta \in \Theta$.

Hemos visto en este capítulo como una familia no necesariamente invariante en sentido clásico lo puede ser en orden respecto a una familia arbitraria de transformaciones biyectivas, bimedibles y conteniendo a la identidad en el espacio muestral. La noción de familia invariante de orden intrínseco α_I engloba la de invariante en sentido clásico, esta sería ahora de orden intrínseco $+\infty$, y va a ser el punto de partida para abordar, en este contexto más amplio, diferentes conceptos y resultados de invariancia clásica.

Capítulo 4

Pérdida invariante en orden

El siguiente elemento a considerar, si seguimos la teoría clásica de invariancia, es la función de pérdida. Si la familia paramétrica de distribuciones de probabilidad en estudio es invariante en sentido clásico, también se le exige a la función de pérdida que lo sea. Si escogemos convenientemente la pérdida, veremos que bajo la perspectiva de la invariancia aproximada también se puede afirmar invariancia pero de un cierto orden.

4.1 Pérdida invariante de orden β_{SC} y β_V

Supongamos que disponemos de una familia paramétrica de distribuciones de probabilidad $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ en un espacio medible (Ω, \mathcal{A}) que además sean absolutamente continuas respecto a una medida σ -finita de referencia μ en (Ω, \mathcal{A}) . Exigiremos, tal y como lo hicimos en el capítulo anterior, que Ω sea un espacio topológico y que el espacio de parámetros Θ sea una variedad n -dimensional.

Sea $\mathcal{G}_T = \{g_t \mid t \in T\}$, donde T es un intervalo de \mathbb{R} , una familia de transformaciones de Ω en Ω , biyectivas y bimedibles que contenga a la identidad. Denotaremos por e aquel elemento de T de manera que g_e sea la identidad sobre Ω y supondremos que $\lim_{t \rightarrow e} g_t w = w$, $\mu - c.s.$

Sea Δ la variedad ambiente y δ una discrepancia en Δ . En este capítulo impondremos una condición adicional a las transformaciones \mathcal{G}_T ; exigiremos que exista una aplicación $c : \Theta \times \Theta \longrightarrow (0, +\infty)$ continua tal que:

$$\lim_{t \rightarrow e} \frac{\delta(g_t^* \theta_1, \theta_1)}{\delta(g_t^* \theta_2, \theta_2)} = c(\theta_1, \theta_2), \quad \forall \theta_1, \theta_2 \in \Theta \quad (4.1)$$

En otras palabras, perturbaciones realizadas sobre dos elementos cualquiera del espacio de parámetros siempre son comparables en el límite. Es una

condición razonable que se verifica en gran cantidad de situaciones y si no se ha exigido anteriormente es porque los resultados que hemos obtenido no la requerían de forma explícita como hipótesis. Verifiquemos que en los ejemplos considerados en el capítulo anterior se cumple.

Ejemplo 4.1.1. Recordemos el ejemplo 3.4.2 donde para la familia paramétrica $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$, el grupo multiplicativo de los reales positivos y discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$ la distancia de Rao al cuadrado, habíamos visto que:

$$\rho^2((a\mu, a), (\mu, 1)) = (\mu^2 + 2)(a - 1)^2 + O((a - 1)^3)$$

De aquí se obtiene que:

$$\lim_{a \rightarrow 1} \frac{\rho^2((a\mu_1, a), (\mu_1, 1))}{\rho^2((a\mu_2, a), (\mu_2, 1))} = \frac{(\mu_1^2 + 2)}{(\mu_2^2 + 2)} > 0, \quad \forall \mu_1, \mu_2 \in \mathbb{R}$$

tal y como queríamos verificar.

Ejemplo 4.1.2. Respecto al ejemplo 3.4.4, en el que recordemos se consideraba la familia de distribuciones de probabilidad $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$, el grupo aditivo real y se tomaba como discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$ la distancia de Rao al cuadrado, teníamos que:

$$\rho^2((a, \sigma), (0, \sigma)) = \frac{a^2}{\sigma^2}$$

De esta manera:

$$\lim_{a \rightarrow 0} \frac{\rho^2((a, \sigma_1), (0, \sigma_1))}{\rho^2((a, \sigma_2), (0, \sigma_2))} = \frac{\sigma_1^2}{\sigma_2^2} > 0, \quad \forall \sigma_1, \sigma_2 \in \mathbb{R}_+$$

con lo que la condición (4.1) también se cumple en este caso.

El objetivo de este capítulo es estudiar el orden de invariancia de la función de pérdida. Pasamos a recordar su definición y algún otro aspecto relacionado. Supondremos que hemos establecido, para el estudio de un cierto fenómeno, un modelo estadístico paramétrico y no entraremos en detalle, al ser un tema que escapa del ámbito de estudio de este trabajo, sobre cómo ha sido seleccionado. Podemos interpretar entonces Ω como el espacio de las mediciones del fenómeno, \mathcal{A} como el de las validaciones sobre las mediciones y \mathcal{P}_Θ como la familia de las posibles distribuciones de probabilidad del fenómeno. Debemos ocuparnos ahora de las afirmaciones a realizar sobre el modelo $(\Omega, \mathcal{A}, \mathcal{P}_\Theta)$ una vez observado $w \in \Omega$. Como ejemplo, supongamos que realizamos un

contraste de hipótesis de la forma $H_0 : \theta \in \Theta_0$ donde $\Theta_0 \subset \Theta$. Podemos tomar dos decisiones: rechazar H_0 o bien no rechazarla. En el caso de la estimación puntual también estamos tomando una decisión al respecto del modelo $(\Omega, \mathcal{A}, \mathcal{P}_\Theta)$ pues queremos estimar el "verdadero" (o una función de éste) parámetro θ . Aquí el espacio de decisiones sería el propio Θ .

De esta manera, cualquier procedimiento para el que a partir de n observaciones se obtiene una decisión indica la existencia de una regla de decisión, concepto que pasamos ahora a definir, pero que debe entenderse como una aplicación del espacio muestral en el conjunto de todas las posibles decisiones. Habitualmente exigiremos que este conjunto sea un espacio en el sentido de la siguiente definición:

Definición 4.1.1. Un *espacio de decisiones* es un espacio medible (D, \mathcal{D}) . Los elementos de D se llaman *decisiones*.

Definición 4.1.2. Sea $(\Omega, \mathcal{A}, \mathcal{P}_\Theta)$ un modelo estadístico paramétrico y consideremos un espacio de decisiones (D, \mathcal{D}) . Una *regla de decisión pura* es una aplicación medible \mathbf{d} de (Ω, \mathcal{A}) en (D, \mathcal{D}) .

Para poder comparar decisiones debemos introducir la siguiente definición.

Definición 4.1.3. Una *función de pérdida* correspondiente al modelo estadístico paramétrico $(\Omega, \mathcal{A}, \mathcal{P}_\Theta)$ y al espacio de decisiones (D, \mathcal{D}) es una aplicación:

$$L : \Theta \times \mathcal{D} \longrightarrow \mathbb{R}_+$$

de manera que $\forall \theta \in \Theta$ la aplicación $L_\theta : d \mapsto L(\theta, d)$ es medible.

De esta manera, dado $\theta \in \Theta$ y $d \in D$ el valor $L(\theta, d)$ puede entenderse como la penalización asociada a escoger d cuando el "verdadero" valor del parámetro es θ . Esto nos lleva a introducir:

Definición 4.1.4. Diremos que d es una *decisión correcta* para el parámetro $\theta \in \Theta$ si y sólo si $L(\theta, d) = 0$.

Supondremos que para cada valor del parámetro θ existe una única decisión correcta d y que cada decisión d es correcta al menos para algún $\theta \in \Theta$. Observemos que estas condiciones se presentan en el caso de la estimación puntual, en el de los test de hipótesis y en el de la determinación de regiones de confianza.

En este trabajo vamos a interesarnos en estimación puntual de los parámetros del modelo. En este caso L es una aplicación:

$$L : \Theta \times \Theta \longrightarrow \mathbb{R}_+$$

en la que $L(\theta', \theta)$ indicará ahora la penalización asociada a tomar como valor del parámetro θ' si en realidad el "verdadero" valor es θ .

Con todos estos conceptos recordados, podemos pasar a analizar la invariancia, en orden, de la pérdida.

Definición 4.1.5. Diremos que L es *invariante de orden escalar* β_{SC} respecto a la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\beta_{SC} = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{(L(\bar{g}_t\theta_1, \bar{g}_t\theta_2) - L(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^*\theta_1, \theta_1)^\beta, \rho^2(g_t^*\theta_2, \theta_2)^\beta)} = 0, \forall \theta_1, \theta_2 \in \Theta\}$$

Definición 4.1.6. Diremos que L es *invariante de orden vectorial* β_V respecto a la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\begin{aligned} \beta_V = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{\tau_{\theta_1}^{\bar{g}_t\theta_1} \text{grad}L(\cdot, \bar{g}_t\theta_2)|_{\bar{g}_t\theta_1}}{\min(\rho^2(g_t^*\theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^*\theta_2, \theta_2)^{\frac{\beta}{2}})} - \\ - \frac{\text{grad}L(\cdot, \theta_2)|_{\theta_1}}{\min(\rho^2(g_t^*\theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^*\theta_2, \theta_2)^{\frac{\beta}{2}})} = 0_{\theta_1}, \forall \theta_1, \theta_2 \in \Theta\} \end{aligned}$$

donde $0_{\theta_1} \in T_{\theta_1}(\Theta)$, $\text{grad}L(\cdot, \theta_2)|_{\theta_1}$ indica el gradiente de la función $L(\cdot, \theta_2)$ en el punto θ_1 y:

$$\tau_{\theta_1}^{\bar{g}_t\theta_1} : T_{\bar{g}_t\theta_1}(\Theta) \longrightarrow T_{\theta_1}(\Theta)$$

es el transporte paralelo desde $T_{\bar{g}_t\theta_1}(\Theta)$ a $T_{\theta_1}(\Theta)$ a lo largo de una geodésica que una $\bar{g}_t\theta_1$ y θ_1 .

Es de esperar que para poder relacionar los órdenes de invariancia de una familia y el de la pérdida que escojamos, la discrepancia y la función de pérdida deberían tener a su vez algún tipo de relación. Recordemos que si la variedad ampliada Δ es Riemanniana, siendo la métrica ρ la informativa, y Θ , el espacio de parámetros, es una subvariedad Riemanniana de Δ de dimensión finita, ha quedado establecida en el capítulo anterior la conveniencia de tomar ρ^2 como medida de discrepancia. El interés en la distancia de Rao como función de pérdida ha sido establecido en Oller y Corcuera [31]. La siguiente sección además recuerda que ρ^2 es invariante en sentido clásico frente a la acción de un grupo. Nos proponemos, en la sección 4.3, reproducir esta propiedad pero en orden.

4.2 Invariancia clásica de la distancia de Rao

Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia de distribuciones de probabilidad en (Ω, \mathcal{A}) absolutamente continuas respecto μ . Supongamos que el espacio de parámetros Θ tenga estructura de n -variedad Riemanniana siendo la métrica la inducida por la matriz de información de Fisher. En este caso podemos considerar como función de pérdida natural la distancia Riemanniana al cuadrado ρ_Θ^2 entre puntos de la variedad Θ , ver Oller y Corcuera [31] y Corcuera [13], recordando nuevamente que esta elección recoge las propiedades geométricas del modelo inducidas por la propia matriz de información de Fisher y resulta ser una generalización de la distancia de Mahalanobis. Además, en términos de invariancia clásica, ρ_Θ^2 es invariante bajo la acción de un grupo de transformaciones biyectivas y bimedibles, si la familia en estudio lo es. Éste es un resultado conocido y utilizado habitualmente que se basa en la conservación de la Información de Fisher frente a transformaciones suficientes. En este trabajo se ha optado por una prueba directa considerando que su importancia así lo requería.

Proposición 4.2.1. *Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de probabilidad y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en (Ω, \mathcal{A}) absolutamente continuas respecto μ . Supongamos que Θ , el espacio de parámetros, tenga estructura de variedad Riemanniana n -dimensional y que la métrica ρ_Θ en Θ sea la inducida por la matriz de información de Fisher.*

Sea $\Phi^\Omega : \Omega \longrightarrow \Omega$ una transformación biyectiva y bimedible y $\Phi : \Theta \longrightarrow \Theta$ biyectiva, de manera que si X es una variable aleatoria siguiendo la distribución P_θ , entonces $\Phi^\Omega(X) \sim P_{\Phi(\theta)}$. Supongamos, además que $\mu \gg P_{\Phi(\theta)}$, $\forall \theta \in \Theta$.

Entonces:

$$\rho_\Theta^2(\Phi(\theta_1), \Phi(\theta_2)) = \rho_\Theta^2(\theta_1, \theta_2), \quad \forall \theta_1, \theta_2 \in \Theta$$

Demostración. Probaremos que la aplicación $\Phi : \Theta \longrightarrow \Theta$ es una isometría y por tanto conserva las distancias entre puntos de la variedad.

Sea $\theta_0 \in \Theta$ y (U, x) y (V, y) sistemas de coordenadas locales en θ_0 y en $\Phi(\theta_0)$ respectivamente. Definimos un nuevo sistema de coordenadas local en θ_0 a través de:

$$\begin{array}{ccc} z : & U \cap \Phi^{-1}(V) & \longrightarrow & \mathbb{R}^n \\ & \theta & & y(\Phi(\theta)) \end{array}$$

Queremos demostrar que $\forall X, Y \in T_{\theta_0}(\Theta)$:

$$\langle X, Y \rangle_{\theta_0} = \langle (\Phi_*)_{\theta_0}(X), (\Phi_*)_{\theta_0}(Y) \rangle_{\Phi(\theta_0)} \quad (4.2)$$

Por simplicidad de notación, utilizaremos el convenio de sumación de Einstein o de índices repetidos. Θ es una variedad Riemanniana n -dimensional por lo que todos los índices, salvo que se indique de manera expresa lo contrario, variarán entre 1 y n . Así, escribiremos:

$$X = X^i \left(\frac{\partial}{\partial x^i} \right)_{\theta_0}, \quad Y = Y^i \left(\frac{\partial}{\partial x^i} \right)_{\theta_0}$$

Por otra parte, X e Y referidos al sistema de coordenadas z se expresan:

$$X = \overline{X}^i \left(\frac{\partial}{\partial z^i} \right)_{\theta_0}, \quad Y = \overline{Y}^i \left(\frac{\partial}{\partial z^i} \right)_{\theta_0}$$

siendo:

$$\overline{X}^\alpha = \left(\frac{\partial z^\alpha}{\partial x^\beta} \right)_{\theta_0} X^\beta, \quad \overline{Y}^\alpha = \left(\frac{\partial z^\alpha}{\partial x^\beta} \right)_{\theta_0} Y^\beta \quad (4.3)$$

y $\left(\left(\frac{\partial z^\alpha}{\partial x^\beta} \right)_{\theta_0} \right)_{\alpha, \beta=1 \dots n}$ la matriz de derivadas parciales de la aplicación:

$$z \circ x^{-1} = y \circ \Phi \circ x^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

evaluadas en el punto $x(\theta_0)$. De esta manera, si notamos $(g_{ij}(\theta_0))_{i,j=1 \dots n}$ como la matriz de la métrica en θ_0 expresada en el sistema de coordenadas x y $(\overline{g}_{ij}(\theta_0))_{i,j=1 \dots n}$ en el sistema de coordenadas z , se tiene que:

$$\langle X, Y \rangle_{\theta_0} = g_{ij}(\theta_0) X^i Y^j = \overline{g}_{ij}(\theta_0) \overline{X}^i \overline{Y}^j \quad (4.4)$$

cumpléndose la relación:

$$\overline{g}_{\alpha\beta}(\theta_0) = \left(\frac{\partial x^\alpha}{\partial z^\beta} \right)_{\theta_0} \left(\frac{\partial x^\beta}{\partial z^j} \right)_{\theta_0} g_{ij}(\theta_0)$$

y siendo ahora $\left(\left(\frac{\partial x^\alpha}{\partial z^\beta} \right)_{\theta_0} \right)_{\alpha, \beta=1 \dots n}$ la matriz de las derivadas parciales de la aplicación $x \circ z^{-1} = x \circ \Phi^{-1} \circ y^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ evaluadas en el punto $z(\theta_0) = y(\Phi(\theta_0))$.

Estudiemos la parte derecha de la igualdad (4.2). Por simplicidad de las notaciones, indicaremos $(\Phi_*)_{\theta_0}(X) = V$, $(\Phi_*)_{\theta_0}(Y) = W$ y escribiremos:

$$V = V^i \left(\frac{\partial}{\partial y^i} \right)_{\Phi(\theta_0)}, \quad W = W^i \left(\frac{\partial}{\partial y^i} \right)_{\Phi(\theta_0)}$$

Entonces, $\forall g \in \mathcal{C}^\infty(\Phi(\theta_0))$:

$$\begin{aligned}
Vg &= X(g \circ \Phi) = X^i \left(\frac{\partial}{\partial x^i} \right)_{\theta_0} (g \circ \Phi) = \\
&= X^i D_i (g \circ \Phi \circ x^{-1})(x(\theta_0)) = X^i D_i (g \circ y^{-1} \circ y \circ \Phi \circ x^{-1})(x(\theta_0)) = \\
&= X^i D_i (g \circ y^{-1} \circ z \circ x^{-1})(x(\theta_0)) = X^i D_\alpha (g \circ y^{-1}(z(\theta_0))) D_i (\Pi^\alpha \circ z \circ x^{-1})(x(\theta_0)) = \\
&= (X^i D_i (\Pi^\alpha \circ z \circ x^{-1})(x(\theta_0))) D_\alpha (g \circ y^{-1})(y(\Phi(\theta_0))) = \\
&= (X^i D_i (\Pi^\alpha \circ z \circ x^{-1})(x(\theta_0))) \left(\frac{\partial}{\partial y^\alpha} \right)_{\Phi(\theta_0)} g
\end{aligned}$$

esto es:

$$V^\alpha = X^i D_i (\Pi^\alpha \circ z \circ x^{-1})(x(\theta_0)) = X^i \left(\frac{\partial z^\alpha}{\partial x^i} \right)_{\theta_0} = \overline{X^\alpha}$$

utilizándose, en la última igualdad, la primera parte de la expresión (4.3). De manera similar, se probaría que $W^\alpha = \overline{Y^\alpha}$. Entonces:

$$\langle (\Phi_*)_{\theta_0}(X), (\Phi_*)_{\theta_0}(Y) \rangle = \tilde{g}_{ij}(\Phi(\theta_0)) \overline{X^i} \overline{Y^j} \quad (4.5)$$

siendo $(\tilde{g}_{ij}(\Phi(\theta_0)))_{i,j=1\dots n}$ la matriz de la métrica en el punto $\Phi(\theta_0)$ referida al sistema de coordenadas y . Estudiemos los elementos de esta matriz. Intentaremos demostrar que:

$$\overline{g}_{ij}(\theta_0) = \tilde{g}_{ij}(\Phi(\theta_0)) \quad , \quad i, j = 1 \dots n \quad (4.6)$$

y la igualdad (4.2) quedará probada. En efecto. Recordemos que nuestro objetivo era calcular $\langle (\Phi_*)_{\theta_0}(X), (\Phi_*)_{\theta_0}(Y) \rangle_{\Phi(\theta_0)} = \tilde{g}_{ij}(\Phi(\theta_0)) V^i W^j$. Entonces, una vez demostrada la expresión anterior y según (4.4) y (4.5):

$$\langle (\Phi_*)_{\theta_0}(X), (\Phi_*)_{\theta_0}(Y) \rangle_{\Phi(\theta_0)} = \overline{g}_{ij}(\theta_0) \overline{X^i} \overline{Y^j} = g_{ij}(\theta_0) X^i Y^j = \langle X, Y \rangle_{\theta_0}$$

con lo que la condición de isometría para Φ , esto es (4.2), estará demostrada. Para probar (4.6) introducimos $p(\cdot; \theta_0) = \frac{dP_{\theta_0}}{d\mu}$ y $p(\cdot; \Phi(\theta_0)) = \frac{dP_{\Phi(\theta_0)}}{d\mu}$. Estas derivadas de Radon-Nikodym existen al ser P_{θ_0} y $P_{\Phi(\theta_0)}$ absolutamente continuas respecto μ por hipótesis. Entonces:

$$\begin{aligned}
\overline{g}_{ij}(\theta_0) &= E_{\theta_0} \left[\frac{\partial}{\partial z^i} \log p(X; \theta_0) \frac{\partial}{\partial z^j} \log p(X; \theta_0) \right] = \\
&= E_{\Phi(\theta_0)} \left[\frac{\partial}{\partial z^i} \log p(\Phi^\Omega(X); \theta_0) \frac{\partial}{\partial z^j} \log p(\Phi^\Omega(X); \theta_0) \right] \quad (4.7)
\end{aligned}$$

Sabemos que:

$$p(\Phi^\Omega(w); \theta_0) = p(w; \Phi(\Theta_0)), \text{ c.s.}$$

pues si $X \sim P_{\theta_0}$, entonces $\Phi^\Omega(X) \sim P_{\Phi(\theta_0)}$. De esta manera, recuperando la expresión (4.7), podemos escribir:

$$\bar{g}_{ij}(\theta_0) = E_{\Phi(\theta_0)} \left[\frac{\partial}{\partial z^i} \log p(X; \Phi(\theta_0)) \frac{\partial}{\partial z^j} \log p(X; \Phi(\theta_0)) \right] = \quad (4.8)$$

Finalmente observemos que:

$$\begin{aligned} \frac{\partial}{\partial z^i} \log p(\cdot; \Phi(\theta_0)) &= D_i(\log p(\cdot; \Phi \circ z^{-1}))(z(\theta_0)) = \\ &= D_i(\log p(\cdot; \Phi \circ (y \circ \Phi)^{-1}))((y \circ \Phi)(\theta_0)) = \frac{\partial}{\partial y^i} \log p(\cdot; \Phi(\theta_0)) \end{aligned}$$

Así la expresión (4.8) se reescribe:

$$\bar{g}_{ij}(\theta_0) = E_{\Phi(\theta_0)} \left[\frac{\partial}{\partial y^i} \log p(X; \Phi(\theta_0)) \frac{\partial}{\partial y^j} \log p(X; \Phi(\theta_0)) \right] = \tilde{g}_{ij}(\theta_0)$$

que es justamente (4.6) tal y como queríamos ver. □

4.3 Invariancia de orden β_W y β_S de ρ^2 .

Una vez demostrada la invariancia en sentido clásico de la distancia Riemanniana al cuadrado ρ^2 vamos a probar que si tomamos como discrepancia ρ^2 y la familia paramétrica de distribuciones de probabilidad es invariante de orden α_I , entonces ρ^2 restringida al espacio de parámetros es una función de pérdida invariante en sentido escalar por lo menos de orden α_I .

Aunque las definiciones de pérdida invariante ya se introdujeron en la sección 4.1, las recuperaremos y las escribiremos de nuevo para la distancia Riemanniana.

Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, Θ una n -variedad y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en (Ω, \mathcal{A}) . Supongamos que Ω es un espacio topológico y que $\mathcal{G}_T = \{g_t \mid t \in T\}$, T un intervalo de \mathbb{R} , es una familia de transformaciones de Ω en Ω , biyectivas y bimedibles conteniendo a la identidad. Sea e el elemento de T de manera que la aplicación g_e es la identidad sobre Ω e impondremos que $\lim_{t \rightarrow e} g_t w = w$ $\mu - c.s.$

Vamos a suponer que la variedad ambiente Δ es una variedad Riemanniana, que su métrica es la informativa y que Θ es una subvariedad Riemanniana de Δ . Tomaremos como discrepancia en Δ la distancia Riemanniana al cuadrado, ρ^2 . Finalmente la distancia Riemanniana en Θ inducida por ρ la indicaremos como ρ_Θ .

La condición (4.1) se escribirá ahora:

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \theta_1, \theta_1)}{\rho^2(g_t^* \theta_2, \theta_2)} = c(\theta_1, \theta_2), \quad \forall \theta_1, \theta_2 \in \Theta \quad (4.9)$$

y las definiciones de pérdida invariante para ρ_Θ^2 serán las siguientes.

Definición 4.3.1. Diremos que ρ_Θ^2 es *invariante de orden escalar* β_{SC} respecto la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\beta_{SC} = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{(\rho_\Theta^2(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - \rho_\Theta^2(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} = 0, \quad \forall \theta_1, \theta_2 \in \Theta\}$$

Definición 4.3.2. Diremos que ρ_Θ^2 es *invariante de orden vectorial* β_V respecto la familia de transformaciones \mathcal{G}_T si y sólo si:

$$\beta_V = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{\tau_{\theta_1}^{\bar{g}_t \theta_1} \exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2) - \exp_{\theta_1}^{-1}(\theta_2)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^* \theta_2, \theta_2)^{\frac{\beta}{2}})} = 0_{\theta_1}, \quad \forall \theta_1, \theta_2 \in \Theta\}$$

donde $0_{\theta_1} \in T_{\theta_1}(\Theta)$ y $\|\exp_{\theta_1}(\theta_2)\|^2 = \rho_\Theta^2(\theta_1, \theta_2)$.

En el caso de la distancia Riemanniana al cuadrado, se pueden relacionar los órdenes de invariancia escalar y vectorial según se muestra en el siguiente resultado.

Proposición 4.3.1. *Los órdenes de invariancia escalar y vectorial de ρ_Θ^2 cumplen la siguiente relación:*

$$\beta_V \leq \beta_{SC}$$

Demostración Sea $\beta \geq 0$ tal que $\forall \theta_1, \theta_2 \in \Theta$:

$$\lim_{t \rightarrow e} \frac{\tau_{\theta_1}^{\bar{g}_t \theta_1} \exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2) - \exp_{\theta_1}^{-1}(\theta_2)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^* \theta_2, \theta_2)^{\frac{\beta}{2}})} = 0 \quad (4.10)$$

Demostraremos que:

$$\lim_{t \rightarrow e} \frac{(\rho_\Theta^2(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - \rho_\Theta^2(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} = 0 \quad (4.11)$$

Esta condición implicará que $\beta \leq \beta_{SC}$ y por ser el supremo de un conjunto la menor de sus cotas superiores tendrá que cumplirse:

$$\beta_V \leq \beta_{SC}$$

tal y como queríamos ver.

Probemos que el límite de (4.11) es efectivamente 0. Si $\theta_1 = \theta_2$ el límite es cero trivialmente. En caso contrario introducimos:

$$A(t) = \tau_{\theta_1}^{\bar{g}_t \theta_1} \exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2)$$

$$B = \exp_{\theta_1}^{-1}(\theta_2)$$

Entonces:

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - \rho_{\Theta}^2(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} = \\ &= \lim_{t \rightarrow e} \frac{(\|A(t)\|^2 - \|B\|^2)^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} = \\ &= \lim_{t \rightarrow e} \frac{(< A(t) + B, A(t) - B >)^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} \leq \\ &\leq \lim_{t \rightarrow e} \frac{\|A(t) - B\|^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} \|A(t) + B\|^2 = 0 \end{aligned}$$

pues:

$$\begin{aligned} &\lim_{t \rightarrow e} \frac{\|A(t) - B\|^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta)} = \\ &= \lim_{t \rightarrow e} \left(\frac{\|\tau_{\theta_1}^{\bar{g}_t \theta_1} \exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2) - \exp_{\theta_1}^{-1}(\theta_2)\|}{\min(\rho^2(g_t^* \theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^* \theta_2, \theta_2)^{\frac{\beta}{2}})} \right)^2 = 0 \end{aligned}$$

según la ecuación (4.10) y al ser $\lim_{t \rightarrow e} \|A(t) + B\|^2 = 4 \rho_{\Theta}^2(\theta_1, \theta_2) \neq 0$ por hipótesis.

De esta manera, la igualdad (4.11) queda probada y la proposición también. \square

Observemos que el resultado anterior no tiene porque haber una igualdad entre β_{SC} y β_V pues podría ser que $\|\exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2)\|^2 = \rho_{\Theta}^2(\theta_1, \theta_2)$, $\forall t \in T$, $\forall \theta_1, \theta_2 \in \Theta$ sin que esto implique necesariamente que:

$$\lim_{t \rightarrow e} \frac{\tau_{\theta_1}^{\bar{g}_t \theta_1} \exp_{\bar{g}_t \theta_1}^{-1}(\bar{g}_t \theta_2) - \exp_{\theta_1}^{-1}(\theta_2)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^{\frac{\beta}{2}}, \rho^2(g_t^* \theta_2, \theta_2)^{\frac{\beta}{2}})} = 0_{\theta_1}$$

En la proposición anterior se ha demostrado la relación entre los órdenes de invariancia escalar y vectorial de ρ_Θ^2 . Observemos que el orden escalar informa sobre la magnitud de la diferencia entre $\rho_\Theta^2(\bar{g}_t\theta_1, \bar{g}_t\theta_2)$ y $\rho_\Theta^2(\theta_1, \theta_2)$ mientras que el orden vectorial aporta además información direccional. La figura 4.1 puede aclarar estas ideas.

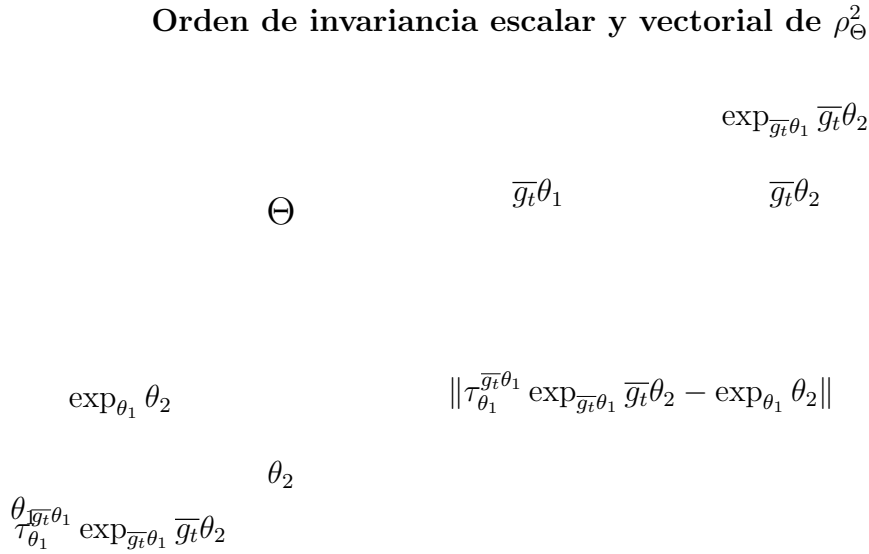


Figura 4.1: Representación gráfica conjunta de $\rho_\Theta^2(\bar{g}_t\theta_1, \bar{g}_t\theta_2)$ y $\rho_\Theta^2(\theta_1, \theta_2)$ frente a $\tau_{\theta_1}^{\bar{g}_t\theta_1} \exp_{\bar{g}_t\theta_1}^{-1} \bar{g}_t\theta_2$ y $\exp_{\theta_1}^{-1} \theta_2$.

De esta manera, nos referiremos al orden escalar de invariancia de ρ_Θ^2 como *orden débil* y al vectorial como *orden fuerte*. Los denotaremos como β_W y β_S respectivamente.

Cabe preguntarse ahora por el orden de invariancia de la función de pérdida en relación al orden de invariancia de la familia. Tal y como hemos visto en la sección 4.2, si la familia de probabilidades es invariante en sentido clásico frente a la acción de un grupo, la distancia de Rao al cuadrado, como función de pérdida, también lo es. Para la situación de invariancia en orden se dispone del siguiente resultado.

Proposición 4.3.2. *Con las notaciones habituales, sea $(\Omega, \mathcal{A}, \mu)$ un espacio de probabilidad y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en (Ω, \mathcal{A}) , absolutamente continuas respecto μ .*

Sea $T \subset \mathbb{R}$ y $\mathcal{G}_T = \{g_t \mid t \in T\}$ una familia de transformaciones biyectivas y bimedibles de Ω en Ω conteniendo a la identidad. Denotemos $e \in T$ aquel elemento para el que g_e sea la identidad sobre Ω .

Supongamos que la familia \mathcal{P}_Θ sea invariante de orden α_I respecto a \mathcal{G}_T y que para t próximos a e el jacobiano de la aplicación \bar{g}_t evaluado en el punto θ sea diferente de 0.

Sea Δ la variedad ampliada. Supondremos que Δ tiene estructura de m -variedad Riemanniana, que la distancia ρ en Δ es la inducida por información de Fisher y que Θ es una subvariedad completa de Δ de dimensión n . Indiquemos por ρ_Θ la distancia en Θ inducida a través de la métrica en Δ . Entonces, ρ_Θ^2 es una función de pérdida invariante en sentido débil de orden $\beta_W \geq \alpha_I$.

Demostración. Sean θ_1 y $\theta_2 \in \Theta$ arbitrarios pero fijos. Vamos a demostrar que si $\alpha < \alpha_I$, entonces:

$$\lim_{t \rightarrow e} \frac{(\rho_\Theta^2(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - \rho_\Theta^2(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0 \quad (4.12)$$

Esta condición implicará que $\alpha \leq \beta_W$ de donde, recordando la definición de supremo de un conjunto, se tendrá que $\alpha_I \leq \beta_W$.

Observemos que si $\theta_1 = \theta_2$ el límite anterior es claramente nulo. Consideremos entonces $\theta_1 \neq \theta_2$.

Antes de abordar la evaluación del límite de (4.12), debemos realizar una serie de cálculos previos que reproducen, en la medida de lo posible, los realizados en la proposición 4.2.1.

Sea $\theta \in \Theta$. Consideremos un sistema de coordenadas local (U, x) en θ que esté adaptado a la subvariedad, esto es, una aplicación:

$$\begin{aligned} x : U &\longrightarrow \mathbb{R}^m \\ p &\longmapsto (x^1(p), \dots, x^m(p)) \end{aligned}$$

de manera que $p \in U \cap \Theta \Leftrightarrow x^1(p) = \dots = x^m(p) = 0$. Un sistema de coordenadas adaptado siempre puede considerarse, tal y como se muestra en las páginas 23 – 24 de Helgason [22].

Sean ahora t próximo a e y U'_t un abierto de Δ tales que:

$$\bar{g}_t : U \cap \Theta \longrightarrow U'_t \cap \Theta$$

sea biyectiva. Consideremos una aplicación:

$$y_t : U'_t \longrightarrow \mathbb{R}^m$$

de manera que (U'_t, y_t) sea un sistema de coordenadas local en $g_t^*\theta$ adaptado a Θ .

Denotemos $\bar{U} = U \cap \Theta$ y $\bar{U}'_t = U'_t \cap \Theta$ e indiquemos, con abuso de notación $x \equiv x|_{\bar{U}}$ e $y_t \equiv y_t|_{\bar{U}'_t}$.

Definimos un sistema de coordenadas local en θ a través de:

$$\begin{array}{ccc} z_t : & U \cap \bar{g}_t^{-1}(\bar{U}'_t) & \longrightarrow \quad \mathbb{R}^n \\ & \vartheta & y_t(\bar{g}_t \vartheta) \end{array}$$

$(U \cap \bar{g}_t^{-1}(\bar{U}'_t), z_t)$ es una carta local de Θ en el punto θ . Completamos z_t a un sistema de coordenadas de Δ que sea adaptado a Θ y cuya restricción a Θ coincida con z_t . Con abuso de notación, lo denotaremos también como z_t . Consideremos U''_t entorno abierto de θ en Δ de manera que $U''_t \cap \Theta = U \cap \bar{g}_t^{-1}(\bar{U}'_t)$. Sin pérdida de generalidad, podemos considerar que $U''_t = U$ pues en caso de que $U''_t \neq U$, tomaríamos en el sistema (U, x) y en el (U''_t, z) con la intersección de U y U''_t . Observemos que, con estas consideraciones, U pasa a depender también de t . Lo denotaremos U_t .

Sean $X, Y \in T_\theta(\Theta)$. Vamos a analizar la diferencia entre las cantidades:

$$\langle X, Y \rangle_\theta \text{ y } \langle (\bar{g}_{t*})_\theta(X), (\bar{g}_{t*})_\theta(Y) \rangle_{\bar{g}_t\theta} \quad (4.13)$$

pues ésta nos indicará lo alejados que nos encontramos de la situación de invariancia clásica, donde $g_t^* = \bar{g}_t$ es una isometría.

Por simplicidad de notación, utilizaremos el convenio de sumación de Einstein o de índices repetidos. Así escribiremos:

$$X = X^i \left(\frac{\partial}{\partial x^i} \right)_{\theta_0}, \quad Y = Y^i \left(\frac{\partial}{\partial x^i} \right)_{\theta_0}$$

siendo $X^i = Y^i = 0$ para $i \in \{n+1, \dots, m\}$, pues recordemos que la carta (U, x) está adaptada a Θ .

Por otra parte, X e Y referidos al sistema de coordenadas z_t se expresan:

$$X = \bar{X}^i \left(\frac{\partial}{\partial z_t^i} \right)_\theta, \quad Y = \bar{Y}^i \left(\frac{\partial}{\partial z_t^i} \right)_\theta$$

siendo:

$$\bar{X}^\alpha = \left(\frac{\partial z_t^\alpha}{\partial x^\beta} \right)_\theta X^\beta, \quad \bar{Y}^\alpha = \left(\frac{\partial z_t^\alpha}{\partial x^\beta} \right)_\theta Y^\beta \quad (4.14)$$

y $((\frac{\partial z_t^\alpha}{\partial x^\beta})_\theta)_{\alpha, \beta=1 \dots n}$ la matriz de derivadas parciales de la aplicación:

$$z_t \circ x^{-1} = y_t \circ \bar{g}_t \circ x^{-1} : x(U_t) \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

evaluadas en el punto $x(\theta)$. Para $i \in \{n+1, \dots, m\}$ sabemos que $X^i = Y^i = 0$ pues recordemos que z_t también está adaptado a Θ .

Sea $(g_{ij}(\theta))_{i,j=1\dots n}$ la matriz de la métrica en θ expresada en el sistema de coordenadas x y $(\bar{g}_{ij}(\theta))_{i,j=1\dots n}$ en el sistema de coordenadas z_t . Podemos pasar de una a la otra a través de:

$$\bar{g}_{\alpha\beta}(\theta_0) = \left(\frac{\partial x^\alpha}{\partial z_t^i}\right)_{\theta_0} \left(\frac{\partial x^\beta}{\partial z_t^j}\right)_{\theta_0} g_{ij}(\theta_0)$$

siendo ahora $((\frac{\partial x^\alpha}{\partial z_t^\beta})_{\theta_0})_{\alpha,\beta=1\dots n}$ la matriz de las derivadas parciales de la aplicación $x \circ z_t^{-1} : (y_t \circ \bar{g}_t)(U_t \cap \bar{g}_t^{-1}(\bar{U}'_t)) \subset \mathbb{R}^m \longrightarrow \mathbb{R}^m$ evaluadas en el punto $z_t(\theta) = y_t^*(\bar{g}_t\theta)$.

Se tiene que:

$$\langle X, Y \rangle_\theta = g_{ij}(\theta_0) X^i Y^j = \bar{g}_{ij}(\theta) \bar{X}^i \bar{Y}^j \quad (4.15)$$

pues el producto escalar entre dos vectores tangentes no depende del sistema de coordenadas escogido para su evaluación.

Estudiemos ahora $\langle (\bar{g}_t)_\theta(X), (\bar{g}_t)_\theta(Y) \rangle_{\bar{g}_t\theta}$. Por simplicidad de notación, indicaremos $(\bar{g}_t)_\theta(X) = V$, $(\bar{g}_t)_\theta(Y) = W$ y escribiremos:

$$V = V^i \left(\frac{\partial}{\partial y_t^i}\right)_{\bar{g}_t\theta}, \quad W = W^i \left(\frac{\partial}{\partial y_t^i}\right)_{\bar{g}_t\theta}$$

Vamos a intentar relacionar V^i con \bar{X}^i y W^i con \bar{Y}^i . Sea $\phi \in \mathcal{C}^\infty(\bar{g}_t\theta)$ y denotemos por Π^α , $\alpha = 1, \dots, m$, las funciones de proyección:

$$\begin{aligned} \Pi^\alpha : \quad \mathbb{R}^m &\longrightarrow \mathbb{R} \\ (p^1, \dots, p^m) &\longmapsto p^\alpha \end{aligned}$$

Se verifica:

$$\begin{aligned} V\phi &= X(\phi \circ \bar{g}_t) = X^i \left(\frac{\partial}{\partial x^i}\right)_\theta (\phi \circ \bar{g}_t) = \\ &= X^i D_i(\phi \circ \bar{g}_t \circ x^{-1})(x(\theta)) = X^i D_i(\phi \circ y_t^{-1} \circ y_t \circ \bar{g}_t \circ x^{-1})(x(\theta)) = \\ &= X^i D_i(\phi \circ y_t^{-1} \circ z_t \circ x^{-1})(x(\theta)) = X^i D_\alpha(\phi \circ y_t^{-1}(z_t(\theta))) D_i(\Pi^\alpha \circ z_t \circ x^{-1})(x(\theta)) = \\ &= \left(X^i D_i(\Pi^\alpha \circ z_t \circ x^{-1})(x(\theta))\right) D_\alpha(\phi \circ y_t^{-1})(y_t(\bar{g}_t\theta)) = \\ &= \left(X^i D_i(\Pi^\alpha \circ z_t \circ x^{-1})(x(\theta))\right) \left(\frac{\partial}{\partial y_t^\alpha}\right)_{\bar{g}_t\theta} \phi \end{aligned}$$

esto es:

$$V^\alpha = X^i D_i (\Pi^\alpha \circ z_t \circ x^{-1})(x(\theta)) = X^i \left(\frac{\partial z_t^\alpha}{\partial x^i} \right)_\theta = \overline{X^\alpha}$$

utilizándose, en la última igualdad, la primera parte de la expresión (4.14). De manera similar, se probaría que $W^\alpha = \overline{Y^\alpha}$. Entonces:

$$\langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t}\theta} = \tilde{g}_{ij}(\overline{g_t}\theta) \overline{X^i} \overline{Y^j}$$

siendo $(\tilde{g}_{ij}(\overline{g_t}\theta))_{i,j=1\dots m}$ la matriz de la métrica en el punto $\overline{g_t}\theta$ referida al sistema de coordenadas y_t . Estudiemos los elementos de esta matriz.

Sea $(\tilde{g}_{ij}(g_t^*\theta))_{i,j=1\dots m}$ la matriz de la métrica en $g_t^*\theta$ referida al sistema de coordenadas y_t . Sabemos que:

$$\tilde{g}_{ij}(g_t^*\theta) = \overline{g}_{ij}(\theta), \quad i, j = 1, \dots, m$$

pues en la proposición 4.2.1 así lo hemos demostrado (recordemos que en la variedad ambiente la transformación $\theta \mapsto g_t^*\theta$ conserva la información de Fisher).

Si recopilamos ahora toda la información de que disponemos, podemos afirmar que:

$$\langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t}\theta} = \tilde{g}_{ij}(\overline{g_t}\theta) \overline{X^i} \overline{Y^j}$$

$$\langle X, Y \rangle_\theta = g_{ij}(\theta_0) X^i Y^j = \tilde{g}_{ij}(g_t^*\theta) \overline{X^i} \overline{Y^j}$$

De esta manera, para comparar $\langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t}\theta}$ con $\langle X, Y \rangle_\theta$ debemos estudiar como difieren los elementos de $(\tilde{g}_{ij}(\overline{g_t}\theta))_{i,j=1,\dots,n}$ de los de la matriz $(\tilde{g}_{ij}(g_t^*\theta))_{i,j=1,\dots,n}$.

Debido a las propiedades del tensor métrico, $\forall i, j \in \{1, \dots, m\}$ la aplicación:

$$\begin{aligned} U'_t &\longrightarrow [0, +\infty) \\ q &\longmapsto \tilde{g}_{ij}(q) \end{aligned}$$

es de clase \mathcal{C}^∞ en U'_t . Por lo tanto, si recordamos la definición de función de clase \mathcal{C}^∞ y componemos con el sistema de coordenadas, la aplicación:

$$\begin{aligned} f_{ij} : y_t(U'_t) \subset \mathbb{R}^m &\longrightarrow U'_t &\longrightarrow [0, +\infty) \\ a &\longmapsto y_t^{-1}(a) &\longmapsto \tilde{g}_{ij}(y_t^{-1}(a)) \end{aligned}$$

también es de clase \mathcal{C}^∞ . Observemos que $y_t(U'_t)$ es un abierto de \mathbb{R}^m al que pertenece $y_t(g_t^*\theta)$. De esta manera, existe $R > 0$ tal que $B(y_t(g_t^*\theta), R) \subset y_t(U'_t)$. Sea:

$$A_t = \sup_{a \in \overline{B(y(g_t^* \theta), R)}} \|\nabla f_{ij}(a)\|$$

donde $\|\cdot\|$ indica la norma euclídea ordinaria en \mathbb{R}^n , esto es:

$$\|v\| = \sqrt{\sum_{i=1}^n v_i^2}, \quad v = (v_1, \dots, v_n) \in \mathbb{R}^n$$

Observemos que el supremo anterior existe al ser $\overline{B(y_t(g_t^* \theta), R)}$ un compacto de \mathbb{R}^m .

Escojamos $r < R$ de manera que $\overline{g_t} \theta \in y_t^{-1}(B(y_t(g_t^* \theta), R))$. La existencia de este $r > 0$ está garantizada si tomamos t próximo a e , pues en este caso $\overline{g_t} \theta$ y $g_t^* \theta$ están arbitrariamente cerca uno del otro ya que las aplicaciones g_t^* y $\overline{g_t}$ tienden a la identidad cuando $t \rightarrow e$.

La aplicación:

$$\begin{array}{ccccc} f_{ij} : B(y_t(g_t^* \theta), R) & \longrightarrow & y_t^{-1}(B(y_t(g_t^* \theta), R)) \subset U'_t & \longrightarrow & [0, +\infty) \\ a & \longmapsto & y_t^{-1}(a) & \longmapsto & \tilde{g}_{ij}(y_t^{-1}(a)) \end{array}$$

es también de clase \mathcal{C}^∞ , pero las derivadas parciales de f_{ij} en $B(y_t(g_t^* \theta), R)$ están acotadas por A_t . Podremos aplicar el Teorema del Valor Medio para funciones de \mathbb{R}^m en \mathbb{R} (que puede encontrarse, entre muchos otros, en Apostol [4], páginas 431–432) y afirmar que:

$$|\tilde{g}_{ij}(g_t^* \theta) - \tilde{g}_{ij}(\overline{g_t} \theta)| \leq A_t \|y_t(g_t^* \theta) - y_t(\overline{g_t} \theta)\|$$

Así:

$$\begin{aligned} & |\langle X, Y \rangle_\theta - \langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t} \theta}| \leq \\ & \leq |\tilde{g}_{ij}(g_t^* \theta) - \tilde{g}_{ij}(\overline{g_t} \theta)| |\overline{X^i} \overline{Y^j}| \leq A_t \|y_t(g_t^* \theta) - y_t(\overline{g_t} \theta)\| |\overline{X^i} \overline{Y^j}| \end{aligned}$$

De aquí obtenemos que:

$$\begin{aligned} \langle X, Y \rangle_\theta^{\frac{1}{2}} & \leq \left(\langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t} \theta} + \right. \\ & \quad \left. + |\langle X, Y \rangle_\theta - \langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t} \theta}| \right)^{\frac{1}{2}} \leq \\ & \leq \left(\langle (\overline{g_{t*}})_\theta(X), (\overline{g_{t*}})_\theta(Y) \rangle_{\overline{g_t} \theta} + A_t \|y_t(g_t^* \theta) - y_t(\overline{g_t} \theta)\| |\overline{X^i} \overline{Y^j}| \right)^{\frac{1}{2}} = \end{aligned}$$

$$= \langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(Y) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} \cdot \left(1 + A_t \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\| \frac{|\overline{X^i} \overline{Y^j}|}{\langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(Y) \rangle_{\overline{g}_t\theta}} \right)^{\frac{1}{2}}$$

Nosotros debemos analizar la situación en la que $X = Y$ pues todo este estudio lo estamos realizando para poder comparar la longitud de una curva σ en Θ con la de las curvas perturbadas por \overline{g}_t y por \overline{g}_t^{-1} , esto es $\overline{g}_t \circ \sigma$ y $\overline{g}_t^{-1} \circ \sigma$.

Observemos que $|\overline{X^i} \overline{X^j}|$ puede entenderse como una norma en $T_{\overline{g}_t\theta}(\Theta)$ (de hecho también en $T_{\overline{g}_t\theta}(\Delta)$ pues $\overline{X^i} = 0$ para $i \in \{n+1, \dots, m\}$). Como en un espacio de dimensión finita todas las normas son equivalentes, $\exists c^1 > 0$, que depende de t y que denotaremos c_t^1 , tal que:

$$\frac{|\overline{X^i} \overline{X^j}|}{\langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}} \leq c_t^1$$

Teniendo en cuenta la observación anterior y recordando (Helgason [22], página 54) que:

$$\lim_{t \rightarrow e} \frac{\|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\|}{\rho(g_t^*\theta, \overline{g}_t\theta)} = 1$$

podemos afirmar que $A_t c_t^1 \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\|$ y $\rho(g_t^*\theta, \overline{g}_t\theta)$ son infinitésimos equivalentes y escribir:

$$\begin{aligned} \langle X, X \rangle_{\theta}^{\frac{1}{2}} &\leq \langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} \cdot \left(1 + A_t c_t^1 \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\| \right)^{\frac{1}{2}} = \\ &= \langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} + \frac{1}{2} A_t c_t^1 \langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} \cdot \\ &\quad \cdot \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\| + O(\rho^2(g_t^*\theta, \overline{g}_t\theta)) \end{aligned} \quad (4.16)$$

De manera similar, probaríamos la existencia de $c_t^2 > 0$ tal que:

$$\begin{aligned} \langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} &\leq \langle X, X \rangle_{\theta}^{\frac{1}{2}} + \frac{1}{2} A_t c_t^2 \langle X, X \rangle_{\theta}^{\frac{1}{2}} \\ &\quad + \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\| + O(\rho^2(g_t^*\theta, \overline{g}_t\theta)) \end{aligned} \quad (4.17)$$

Sea $c_t = \max(c_t^1, c_t^2)$. Recopilando la información de (4.16) y de (4.17), obtenemos:

$$\begin{aligned} &|\langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} - \langle X, X \rangle_{\theta}^{\frac{1}{2}}| \leq \\ &\leq A_t c_t \|y_t(g_t^*\theta) - y_t(\overline{g}_t\theta)\| \left(\langle (\overline{g}_{t*})_\theta(X), (\overline{g}_{t*})_\theta(X) \rangle_{\overline{g}_t\theta}^{\frac{1}{2}} + \langle X, X \rangle_{\theta}^{\frac{1}{2}} \right) + O(\rho^2(g_t^*\theta, \overline{g}_t\theta)) \end{aligned} \quad (4.18)$$

Sean ahora θ_1 y $\theta_2 \in \Theta$, $\theta_1 \neq \theta_2$, y $\sigma : [a, b] \subset \mathbb{R} \longrightarrow \Theta$ una curva en Θ de manera que $\rho_\Theta(\theta_1, \theta_2) = l(\sigma)$. Recordemos que $l(\sigma)$ indica la longitud de σ y que una curva σ verificando la condición anterior siempre existe tal y como puede consultarse en el apéndice de este trabajo o bien en cualquier texto de Geometría Diferencial. Estudiemos como difieren $l(\sigma)$ y $l(\bar{g}_t \circ \sigma)$.

$$\begin{aligned} l(\sigma) &= \int_a^b \|(\sigma_*)_s(\frac{d}{ds})\|_{\sigma(s)} ds \\ l(\bar{g}_t \circ \sigma) &= \int_a^b \|((\bar{g}_t \circ \sigma)_*)(\frac{d}{ds})\|_{\bar{g}_t \sigma(s)} ds = \\ &= \int_a^b \|(\bar{g}_t)_\sigma(s)((\sigma_*)_s(\frac{d}{ds}))\|_{\bar{g}_t \sigma(s)} ds \end{aligned}$$

Denotemos por X el vector tangente $(\sigma_*)_s(\frac{d}{ds}) \in T_{\sigma(s)}(\Theta)$. Debido a (4.18), podemos afirmar que:

$$\begin{aligned} |l(\sigma) - l(\bar{g}_t \circ \sigma)| &= \left| \int_a^b \|X\|_{\sigma(s)} ds - \int_a^b \|(\bar{g}_t)_\sigma(s)(X)\|_{\bar{g}_t \sigma(s)} ds \right| \leq \\ &\leq \int_a^b A_t c_t \|y_t(g_t^* \sigma(s)) - y_t(\bar{g}_t \sigma(s))\| \\ &\quad \left(\|(\bar{g}_t)_\sigma(s)(X)\|_{\bar{g}_t \sigma(s)} + \|X\|_{\sigma(s)} + O(\rho^2(g_t^* \sigma(s), \bar{g}_t \sigma(s))) \right) ds \end{aligned}$$

Repitiendo todo el proceso anterior pero con \bar{g}_t^{-1} en el papel de \bar{g}_t , podemos afirmar que si $\gamma : [a', b'] \subset \mathbb{R} \rightarrow \Theta$ es una curva en Θ de manera que $\rho_\Theta(\bar{g}_t \theta_1, \bar{g}_t \theta_2) = l(\gamma)$, entonces:

$$\begin{aligned} |l(\gamma) - l(\bar{g}_t^{-1} \circ \gamma)| &\leq \int_{a'}^{b'} A_c \|y(g_t^{*-1} \gamma(s)) - y(\bar{g}_t^{-1} \gamma(s))\| \\ &\quad \left(\|(\bar{g}_t^{-1})_\gamma(s)(Y)\|_{\bar{g}_t^{-1} \gamma(s)} + \|Y\|_{\gamma(s)} + O(\rho^2(g_t^{*-1} \gamma(s), \bar{g}_t^{-1} \gamma(s))) \right) ds \end{aligned}$$

Observemos que si Θ no fuese una subvariedad completa de Δ , no podríamos afirmar la existencia de una curva γ de manera que $\rho_\Theta(\bar{g}_t \theta_1, \bar{g}_t \theta_2) = l(\gamma)$. Deberíamos emplear entonces el argumento de existencia de una curva γ_ϵ de manera que $|\rho_\Theta(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - l(\gamma)| < \epsilon$, $\epsilon > 0$, y verificar que los pasos que vamos a realizar a continuación son válidos independientemente del ϵ escogido. Por simplicidad de la demostración se ha optado por exigir la condición de completitud de la subvariedad.

Sean:

$$I_\sigma(t) = \int_a^b Ac \left\| y(g_t^* \sigma(s)) - y(\overline{g}_t \sigma(s)) \right\|$$

$$\left(\left\| (\overline{g}_{t*})_{\sigma(s)}(X) \right\|_{\overline{g}_t \sigma(s)} + \|X\|_{\sigma(s)} + O(\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))) \right) ds$$

$$I_\gamma(t) = \int_{a'}^{b'} Ac \left\| y(g_t^{*-1} \gamma(s)) - y(\overline{g}_t^{-1} \gamma(s)) \right\|$$

$$\left(\left\| (\overline{g}_{t*}^{-1})_{\gamma(s)}(Y) \right\|_{\overline{g}_t^{-1} \gamma(s)} + \|Y\|_{\gamma(s)} + O(\rho^2(g_t^{*-1} \gamma(s), \overline{g}_t^{-1} \gamma(s))) \right) ds$$

Conservando las notaciones anteriores, se tiene:

$$\rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) - \rho_\Theta(\theta_1, \theta_2) = \rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) - l(\sigma) \leq$$

$$\leq l(\overline{g}_t \circ \sigma) - l(\sigma) \leq |l(\overline{g}_t \circ \sigma) - l(\sigma)| \leq A_t c_t I_\sigma(t)$$

De manera similar:

$$\rho_\Theta(\theta_1, \theta_2) - \rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) = \rho_\Theta(\theta_1, \theta_2) - l(\gamma) \leq$$

$$\leq l(\overline{g}_t^{-1} \circ \gamma) - l(\gamma) \leq |l(\overline{g}_t^{-1} \circ \gamma) - l(\gamma)| \leq A_t c_t I_\gamma(t)$$

Por lo tanto:

$$|\rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) - \rho_\Theta(\theta_1, \theta_2)|^2 \leq (A_t c_t)^2 \min(I_\sigma^2(t), I_\gamma^2(t))$$

Pasemos ahora a demostrar que el límite de (4.12) es efectivamente 0. Sea $\alpha < \alpha_I$:

$$\lim_{t \rightarrow e} \frac{(\rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) - \rho_\Theta(\theta_1, \theta_2))^2}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} \leq$$

$$\leq \lim_{t \rightarrow e} (A_t c_t)^2 \frac{\min(I_\sigma^2(t), I_\gamma^2(t))}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} (\rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) + \rho_\Theta(\theta_1, \theta_2))^2$$

Para ver que el límite anterior es nulo, únicamente es necesario probar que:

$$\lim_{t \rightarrow e} \frac{\min(I_\sigma^2(t), I_\gamma^2(t))}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0 \quad (4.19)$$

pues:

$$\lim_{t \rightarrow e} (\rho_\Theta(\overline{g}_t \theta_1, \overline{g}_t \theta_2) + \rho_\Theta(\theta_1, \theta_2))^2 = 4\rho_\Theta(\theta_1, \theta_2) \neq 0$$

Veamos que el límite de (4.19) es 0. Observemos que basta demostrar:

$$\lim_{t \rightarrow e} \frac{I_\sigma^2(t)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0$$

$$\lim_{t \rightarrow e} \frac{I_\gamma^2(t)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0$$

Analicemos en primer lugar el límite correspondiente a I_σ . Según la desigualdad de Hölder, podemos afirmar que:

$$\begin{aligned} & \lim_{t \rightarrow e} \frac{I_\sigma^2(t)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} \leq \\ & \lim_{t \rightarrow e} \int_a^b A_t c_t \left(\|(\overline{g}_t)_\sigma(s)(X)\|_{\overline{g}_t \sigma(s)} + \|X\|_{\sigma(s)} \right)^2 \frac{\|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))} \cdot \\ & \quad \cdot \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} \cdot \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} ds + \\ & \quad + \lim_{t \rightarrow e} \int_a^b A_t c_t \frac{O(\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s)) \|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} ds \end{aligned}$$

Sea L_1 el primero de los dos límites anteriores y L_2 el segundo. Vamos a demostrar que $L_1 = L_2 = 0$.

Sabemos que las distancias Riemanniana y Euclídea siempre son comparables en el límite (recordemos que más detalles pueden consultarse en la página 54 de Helgason [22]). De esta manera, podemos afirmar que:

$$\lim_{t \rightarrow e} \frac{\|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))} = 1, \quad \forall s \in [a, b] \quad (4.20)$$

Por otra parte, la familia \mathcal{P}_Θ es invariante de orden α_I respecto a \mathcal{G}_T . Así, al ser $\alpha < \alpha_I$:

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} = 0 \quad (4.21)$$

Finalmente y debido a la condición adicional que hemos impuesto en este tema a las transformaciones g_t^* :

$$\begin{aligned} \lim_{t \rightarrow e} \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} &= \lim_{t \rightarrow e} \frac{1}{\min\left(\frac{\rho^2(g_t^* \theta_1, \theta_1)^\alpha}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}, \frac{\rho^2(g_t^* \theta_2, \theta_2)^\alpha}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}\right)} = \\ &= \lim_{t \rightarrow e} \frac{1}{\min(c(\theta_1, \sigma(s)), c(\theta_2, \sigma(s)))} \neq 0 \end{aligned} \quad (4.22)$$

Ahora, el Lema de Fatou informa que:

$$\begin{aligned}
& \lim_{t \rightarrow e} \int_a^b A_t c_t \left(\|(\overline{g}_{t*})_{\sigma(s)}(X)\|_{\overline{g}_t \sigma(s)} + \|X\|_{\sigma(s)} \right)^2 \cdot \frac{\|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))} \\
& \quad \cdot \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} \cdot \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} ds \leq \\
& \leq \limsup_{t \rightarrow e} \int_a^b A_t c_t \left(\|(\overline{g}_{t*})_{\sigma(s)}(X)\|_{\overline{g}_t \sigma(s)} + \|X\|_{\sigma(s)} \right)^2 \cdot \frac{\|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))} \\
& \quad \cdot \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} \cdot \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} ds \leq \\
& \leq \int_a^b A_t c_t \limsup_{t \rightarrow e} \left(\|(\overline{g}_{t*})_{\sigma(s)}(X)\|_{\overline{g}_t \sigma(s)} + \|X\|_{\sigma(s)} \right)^2 \cdot \frac{\|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))} \\
& \quad \cdot \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} \cdot \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} ds = 0
\end{aligned}$$

con lo que queda demostrado que $L_1 = 0$. Probemos ahora que $L_2 = 0$. Utilizaremos la misma técnica que para L_1 y por lo tanto basta demostrar que:

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s)) \|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0$$

Pero esta condición se verifica pues:

$$\begin{aligned}
& \lim_{t \rightarrow e} \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s)) \|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\|}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = \\
& = \lim_{t \rightarrow e} \|y_t(g_t^* \sigma(s)) - y_t(\overline{g}_t \sigma(s))\| \cdot \frac{\rho^2(g_t^* \sigma(s), \overline{g}_t \sigma(s))}{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha} \cdot \frac{\rho^2(g_t^* \sigma(s), \sigma(s))^\alpha}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha)} = 0
\end{aligned}$$

donde en la última igualdad se han vuelto a utilizar las ecuaciones (4.21) y (4.22). El lema de Fatou informa que $L_2 = 0$ y podemos afirmar que:

$$\lim_{t \rightarrow e} \frac{I_\sigma^2(t)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0$$

De manera similar se vería que:

$$\lim_{t \rightarrow e} \frac{I_\gamma^2(t)}{\min(\rho^2(g_t^* \theta_1, \theta_1)^\alpha, \rho^2(g_t^* \theta_2, \theta_2)^\alpha)} = 0$$

y por lo tanto queda demostrado que ρ_Θ es invariante por lo menos de orden α_I .

□

4.4 Ejemplos.

Vamos a verificar que la proposición 4.3.2 se cumple en los ejemplos 3.4.2 y 3.4.4 pues para ellos ya tenemos calculados los ordenes de invariancia de las familias.

Ejemplo 4.4.1. Volvamos al ejemplo 3.4.2 con la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$, el grupo multiplicativo de los reales positivos y la distancia de Rao al cuadrado como discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$.

Consultando Burbea [9], tenemos que $\forall \mu_1, \mu_2 \in \mathbb{R}$ y $\forall a \in \mathbb{R}_+$:

$$\left(\rho_{\Theta}^2((a\mu_1, 1), (a\mu_2, 1)) - \rho_{\Theta}^2((\mu_1, 1)(\mu_2, 1)) \right)^2 = ((\mu_1 - \mu_2)^2 (a^2 - 1))^2$$

Por otra parte:

$$\rho^2((a\mu, a), (\mu, 1)) = (\mu^2 + 2)(a - 1)^2 + O((a - 1)^3), \quad \forall \mu \in \mathbb{R}, \quad \forall a \in \mathbb{R}_+$$

De esta manera $\forall \mu \in \mathbb{R}$:

$$\lim_{a \rightarrow 1} \frac{(\rho_{\Theta}^2((a\mu_1, 1), (a\mu_2, 1)) - \rho_{\Theta}^2((\mu_1, 1)(\mu_2, 1)))^2}{\min\{\rho^2((a\mu_1, a), (\mu_2, 1))^\beta, \rho^2((a\mu_2, a), (\mu_2, 1))^\beta\}} = 0 \quad , \beta < 1$$

esto es, la función de pérdida es invariante débil de orden $\beta_W = 1$.

Ejemplo 4.4.2. En el ejemplo 3.4.4 donde se consideraba la familia de distribuciones de probabilidad $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$, el grupo aditivo real y se tomaba como discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$ la distancia de Rao al cuadrado, teníamos que:

$$\rho^2((a, \sigma), (0, \sigma)) = \frac{a^2}{\sigma^2}, \quad \forall \sigma \in \mathbb{R}_+, \quad \forall a \in \mathbb{R}$$

Por otra parte:

$$\begin{aligned} & \rho_{\Theta}^2((0, \sqrt{\sigma_1^2 + \frac{a^2}{2}}), (0, \sqrt{\sigma_2^2 + \frac{a^2}{2}})) - \rho_{\Theta}^2((0, \sigma_1), (0, \sigma_2)) = \\ & = 2(\log(\sqrt{\frac{\sigma_1^2 + \frac{a^2}{2}}{\sigma_2^2 + \frac{a^2}{2}}}))^2 - 2(\log(\frac{\sigma_1}{\sigma_2}))^2 = C \cdot a^2 + O(a^3) \end{aligned}$$

siendo C una constante diferente de 0.

De esta manera, $\forall \sigma_2, \sigma_2 \in \mathbb{R}_+$:

$$\lim_{a \rightarrow 0} \frac{\rho_{\Theta}^2((0, \sqrt{\sigma_1^2 + \frac{a^2}{2}}), (0, \sqrt{\sigma_2^2 + \frac{a^2}{2}})) - \rho_{\Theta}^2((0, \sigma_1), (0, \sigma_2))}{\min\{\rho^2((a, \sigma_1), (0, \sigma_1))^\beta, \rho^2((a, \sigma_2), (0, \sigma_2))^\beta\}} = 0 \quad , \beta < 2$$

esto es, la función de pérdida es invariante débil de orden 2.

Capítulo 5

Estimador equivariante en orden

El tercer elemento que vamos a analizar en este trabajo es el de los estimadores de los parámetros del modelo. Si disponemos de una familia invariante es natural (ver [20] o bien [30]) restringir nuestra atención a la clase de los estimadores equivariantes. Sin embargo, una familia de distribuciones de probabilidad no necesariamente será invariante en sentido clásico tal y como hemos visto en los capítulos anteriores. Y aunque así sea, un estimador puede no verificar la propiedad de equivariancia en sentido clásico y sí serlo en orden, concepto que introduciremos en este capítulo. Una vez que esta definición haya sido establecida, veremos que en nuestra situación de invariancia en orden, además de ser una clase razonable de estimadores, los equivariantes en orden cumplen una serie de propiedades interesantes en cuanto al sesgo intrínseco y al riesgo Riemanniano.

5.1 Estimador equivariante de órdenes γ_W y

γ_S

Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia de distribuciones de probabilidad en (Ω, \mathcal{A}) absolutamente continuas respecto μ . Supondremos que Ω es un espacio topológico y que el espacio de parámetros Θ es una n -variedad de clase \mathcal{C}^∞ .

Sea T un intervalo de \mathbb{R} y $\mathcal{G}_T = \{g_t \mid t \in T\}$ una familia de transformaciones de Ω en Ω , biyectivas y bimedibles conteniendo a la identidad. Sea e el elemento de T de manera que g_e sea la identidad sobre Ω y supongamos que $\lim_{t \rightarrow e} g_t w = w$, *c.s.*

Sea (Δ, ρ) la variedad ambiente que supondremos verificando las condiciones

del ejemplo 3.1.1. Recordemos que en tal situación tomábamos como métrica en Δ la informativa y su restricción a Θ , que denotaremos como ρ_Θ , elevada al cuadrado como función de pérdida. Finalmente hacer notar que hemos impuesto la existencia de una aplicación $c : \Theta \times \Theta \longrightarrow (0, +\infty)$ tal que, para todo elemento de \mathcal{G}_T , se verifique:

$$\lim_{t \rightarrow e} \frac{\rho(g_t^* \theta_1, \theta_1)}{\rho(g_t^* \theta_2, \theta_2)} = c(\theta_1, \theta_2), \quad \forall \theta_1, \theta_2 \in \Theta \quad (5.1)$$

Sea $U : \Omega^k \longrightarrow \Theta$ un estimador de $\theta \in \Theta$ en el modelo estadístico paramétrico $\{(\Omega, \mathcal{A}, P_\theta), \theta \in \Theta\}$. En la situación de invariancia clásica, si \mathcal{P}_Θ es una familia de distribuciones de probabilidad invariante frente a la acción de un grupo $\mathcal{G}_T = \{g_t : \Omega \rightarrow \Omega \mid t \in T\}$, diremos que el estimador U es *equivariante bajo la acción de \mathcal{G}_T* si y sólo si:

$$U(g_t X) = \bar{g}_t U(X)$$

donde $X = (x_1, \dots, x_k) \in \Omega^k$, $g_t X = (g_t x_1, \dots, g_t x_k)$ y \bar{g}_t es la transformación inducida por g_t .

En nuestra situación de invariancia en orden debemos adaptar esta definición convenientemente. Observemos que va a ser necesario introducir una medida de lo alejados que se encuentran los estimadores $U(g_t X)$ y $\bar{g}_t U(X)$. De igual manera a como lo hicimos en el estudio de la función de pérdida, proponemos dos maneras de cuantificar esta diferencia. Más adelante veremos que están relacionadas.

Con las notaciones anteriores, sea $U : \Omega^k \rightarrow \Theta$ un estimador de $\theta \in \Theta$.

Definición 5.1.1. Diremos que U es *equivariante de orden débil γ_W* respecto a la familia \mathcal{G}_T si y sólo si:

$$\gamma_W = \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s. } \forall \theta', \theta \in \Theta\}$$

Definición 5.1.2. Diremos que U es *equivariante de orden fuerte γ_S* respecto a la familia \mathcal{G}_T si y sólo si:

$$\gamma_S = \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s. } \forall \theta', \theta \in \Theta\}$$

Observemos que en la definición anterior podemos tomar $\theta' = U(X)$. Entonces, $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(U(g_t X), \bar{g}_t U(X))}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s. } \forall \theta \in \Theta$$

condición que cabe esperar al generalizar el concepto de estimador equivariante en sentido clásico.

Tal y como hemos anunciado anteriormente, γ_S y γ_W se encuentran relacionados. El siguiente resultado analiza este aspecto.

Proposición 5.1.1. *Los órdenes de equivariancia fuerte y débil de un estimador cumplen la siguiente relación:*

$$\gamma_S \leq \gamma_W$$

Demostración. Sea $\theta', \theta \in \Theta$ y consideremos $\gamma > 0$ tal que:

$$\lim_{t \rightarrow e} \frac{\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s.} \quad (5.2)$$

Veremos que γ también cumple que:

$$\lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(U(g_t X), \theta') - \rho_{\Theta}^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s.} \quad (5.3)$$

Esta condición implicará que $\gamma \leq \gamma_W$ y recordando la definición de supremo de un conjunto, obtendremos que $\gamma_S \leq \gamma_W$.

Así, nuestro objetivo se ha visto reducido a demostrar que el límite de (5.3) es cero. Observemos que, como consecuencia de la desigualdad triangular:

$$| \|\exp_{\theta'}^{-1} U(g_t X)\| - \|\exp_{\theta'}^{-1} \bar{g}_t U(X)\| | \leq \|\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)\|$$

Multiplicando por el conjugado del argumento del valor absoluto y elevando al cuadrado, obtenemos que:

$$\begin{aligned} & (\rho_{\Theta}^2(U(g_t X), \theta') - \rho_{\Theta}^2(\bar{g}_t U(X), \theta'))^2 \leq \\ & \leq \|\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)\|^2 \cdot (\|\exp_{\theta'}^{-1} U(g_t X)\| + \|\exp_{\theta'}^{-1} \bar{g}_t U(X)\|)^2 \end{aligned}$$

Utilizando la desigualdad anterior, podemos afirmar que:

$$0 \leq \lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(U(g_t X), \theta') - \rho_{\Theta}^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq$$

$$\leq \lim_{t \rightarrow e} \frac{\|\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)\|^2}{\rho^2(g_t^* \theta, \theta)^\gamma} \cdot (\|\exp_{\theta'}^{-1} U(g_t X)\| + \|\exp_{\theta'}^{-1} \bar{g}_t U(X)\|)^2 = 0, \text{ c.s.}$$

en virtud de (5.2) y debido a que:

$$\lim_{t \rightarrow e} (\|\exp_{\theta'}^{-1} U(g_t X)\| + \|\exp_{\theta'}^{-1} \bar{g}_t U(X)\|)^2 = 4 \|\exp_{\theta'}^{-1} U(X)\|^2 \neq 0 \text{ c.s.}$$

De esta manera, queda verificado que el límite de (5.3) es cero y la proposición demostrada. \square

5.2 Equivariancia en orden, riesgo y sesgo

En esta sección vamos a analizar como la equivariancia en orden de un estimador U nos informa sobre el comportamiento en términos de sesgo y riesgo de los estimadores perturbados en el espacio muestral $U(g_t \cdot)$ y en el espacio de parámetros $\bar{g}_t U(\cdot)$. En el caso de invariancia clásica uno y otro estimador coinciden. Veamos que podemos afirmar en nuestra situación de invariancia en orden.

Estudiaremos, en primer lugar el riesgo Riemanniano. Recordemos que si U es un estimador de $\theta \in \Theta$, el riesgo Riemanniano (ver capítulo 1) asociado a U es:

$$E_\theta \left(\rho_\Theta^2(U(X), \theta) \right), \forall \theta \in \Theta$$

donde el espacio de parámetros Θ se supone con estructura de variedad Riemanniana siendo ρ_Θ la distancia Riemanniana en Θ .

En nuestra situación, nos interesaría demostrar que, si las perturbaciones de los parámetros del modelo son cercanas a la identidad, el riesgo Riemanniano medio se conserva. Debemos analizar con cautela este requerimiento.

Fijado $t \in T$, y $\theta \in \Theta$ la familia de transformaciones \mathcal{G}_T perturba el valor del parámetro θ transformándolo en $\bar{g}_t \theta$. De esta manera, el nuevo parámetro a estimar es $\bar{g}_t \theta$. Nuestro objetivo sería demostrar que el riesgo Riemanniano asociado a el estimador perturbado no se aleja demasiado del asociado a U . Pero, ¿qué estimador perturbado debemos seleccionar? A priori disponemos de dos candidatos: $U(g_t \cdot)$ y $\bar{g}_t U(\cdot)$. Tal y como muestra el siguiente resultado, si el estimador U es equivariante de orden débil, ambos verifican la condición deseada.

Proposición 5.2.1. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y consideremos una familia de distribuciones de probabilidad en (Ω, \mathcal{A}) , $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$, absolutamente continuas respecto a μ .

Sea $\mathcal{G}_T = \{g_t : \Omega \rightarrow \Omega \mid t \in T\}$, $T \subset \mathbb{R}$, una familia de transformaciones biyectivas y bimedibles conteniendo a la identidad. Sea $e \in T$ aquel elemento de T para el que g_e es la aplicación identidad sobre Ω .

Denotemos por (Δ, ρ) la variedad ampliada y supongamos que ρ_Θ^2 sea una función de pérdida invariante de orden débil β_W .

Sea $U : \Omega \rightarrow \Theta$ un estimador de los parámetros del modelo. Entonces:

$$\beta_W \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))|^2}{\rho^2(g_t^*\theta, \theta)^\gamma} = 0, \forall \theta \in \Theta\}$$

Si además U es un estimador equivariante de orden débil γ_W :

$$\beta_W \wedge \gamma_W \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(U(g_t X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))|^2}{\rho^2(g_t^*\theta, \theta)^\gamma} = 0, \forall \theta \in \Theta\}$$

Demostración. Sea $\theta \in \Theta$. Según la desigualdad de Cauchy–Schwarz, se verifica que:

$$\begin{aligned} |E_\theta(\rho_\Theta^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))|^2 &\leq \\ &\leq E_\theta((\rho_\Theta^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))^2) \end{aligned}$$

Por otra parte, fijado $w \in \Omega$, la aplicación:

$$T \ni t \mapsto \frac{(\rho_\Theta^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))^2}{\rho^2(g_t^*\theta, \theta)^\gamma}$$

es continua. Sin perder generalidad (únicamente estamos interesados en perturbaciones cercanas a la identidad), podemos considerar que T es un compacto de \mathbb{R} que contiene a e . De esta manera, debe existir $t_0 = t_0(w)$ tal que:

$$\frac{(\rho_\Theta^2(\overline{g_{t_0}}U(X), \overline{g_{t_0}}\theta) - \rho_\Theta^2(U(X), \theta))^2}{\rho^2(g_{t_0}^*\theta, \theta)^\gamma} = \sup_{t \in T} \frac{(\rho_\Theta^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_\Theta^2(U(X), \theta))^2}{\rho^2(g_t^*\theta, \theta)^\gamma}$$

Sea Z la variable aleatoria definida según:

$$Z = \frac{(\rho_{\Theta}^2(\overline{g_{t_0}}U(X), \overline{g_{t_0}}\theta) - \rho_{\Theta}^2(U(X), \theta))^2}{\rho^2(g_{t_0}^*\theta, \theta)^\gamma}$$

Observemos que Z es integrable y que domina a $\frac{(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))^2}{\rho^2(g_t^*\theta, \theta)^\gamma}$, $\forall t \in T$. Así, podemos aplicar el Teorema de Convergencia Dominada de Lebesgue y afirmar que, $\forall \gamma < \beta_W$:

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow e} \frac{|E_{\theta}(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))|^2}{\rho^2(g_t^*\theta, \theta)^\gamma} \leq \\ &\leq \lim_{t \rightarrow e} \frac{E_{\theta}((\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))^2)}{\rho^2(g_t^*\theta, \theta)^\gamma} = \\ &= E_{\theta}\left(\lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))^2}{\rho^2(g_t^*\theta, \theta)^\gamma}\right) = 0 \end{aligned}$$

Pues la función de pérdida ρ_{Θ}^2 es invariante de orden débil β_W , con lo que $\forall \gamma < \beta_W$:

$$\lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))^2}{\rho^2(g_t^*\theta, \theta)^\gamma} = 0, \text{ c.s.}$$

Así:

$$\beta_W \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_{\theta}(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))|^2}{\rho^2(g_t^*\theta, \theta)^\gamma} = 0, \forall \theta \in \Theta\}$$

tal y como queríamos ver.

Supongamos ahora que U sea un estimador equivariante de orden débil γ_W .

De esta manera, podemos afirmar que $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(U(g_t X), \theta') - \rho_{\Theta}^2(\overline{g_t} U(X), \theta')}{\rho^2(g_t^*\theta, \theta)^\gamma} = 0, \text{ c.s. } \forall \theta', \theta \in \Theta$$

Por otra parte:

$$\begin{aligned} &|E_{\theta}(\rho_{\Theta}^2(U(g_t X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))|^2 \leq \\ &\leq 2 |E_{\theta}(\rho_{\Theta}^2(U(g_t X), \overline{g_t}\theta) - \rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta))|^2 + \\ &\quad + 2 |E_{\theta}(\rho_{\Theta}^2(\overline{g_t}U(X), \overline{g_t}\theta) - \rho_{\Theta}^2(U(X), \theta))|^2 \end{aligned}$$

Hemos demostrado anteriormente que para todo estimador U se verifica:

$$\lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_\Theta^2(U(X), \theta))|^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \quad \forall \theta \in \Theta$$

siempre que $\gamma < \beta_W$. Vamos a demostrar ahora que:

$$\lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(U(g_t X), \bar{g}_t \theta) - \rho_\Theta^2(\bar{g}_t U(X), \bar{g}_t \theta))|^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \quad \forall \theta \in \Theta$$

si $\gamma < \gamma_W$. De esta manera quedará demostrada la proposición 5.2.1.

Recordemos que hemos supuesto que $T \subset \mathbb{R}$ es un compacto de \mathbb{R} . De esta manera, el conjunto $K = \{\bar{g}_t \theta \mid \theta \in \Theta\}$ también lo es por ser la imagen de un compacto por una aplicación continua. Así, debe existir $\theta' \in K$ de manera que:

$$\frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = \sup_{\tilde{\theta} \in K} \frac{(\rho_\Theta^2(U(g_t X), \tilde{\theta}) - \rho_\Theta^2(\bar{g}_t U(X), \tilde{\theta}))^2}{\rho^2(g_t^* \theta, \theta)^\gamma}$$

Por otra parte, observemos que de igual manera a como lo hemos hecho anteriormente podemos dominar $\frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma}$, $\forall t \in T$, por una variable aleatoria integrable. De esta manera, tiene sentido aplicar el Teorema de Convergencia Dominada de Lebesgue y afirmar que, $\forall \gamma < \gamma_W$:

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(U(g_t X), \bar{g}_t \theta) - \rho_\Theta^2(\bar{g}_t U(X), \bar{g}_t \theta))|^2}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq \\ &\leq \lim_{t \rightarrow e} \frac{E_\theta((\rho_\Theta^2(U(g_t X), \bar{g}_t \theta) - \rho_\Theta^2(\bar{g}_t U(X), \bar{g}_t \theta))^2)}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq \\ &\leq \lim_{t \rightarrow e} E_\theta\left(\frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma}\right) = \\ &= E_\theta\left(\lim_{t \rightarrow e} \frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma}\right) = 0 \end{aligned}$$

pues recordemos que por ser $\gamma < \gamma_W$:

$$\lim_{t \rightarrow e} \frac{(\rho_\Theta^2(U(g_t X), \theta') - \rho_\Theta^2(\bar{g}_t U(X), \theta'))^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \quad c.s. \quad \forall \theta', \theta \in \Theta$$

Queda así demostrado que:

$$\beta_W \wedge \gamma_W \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_\theta(\rho_\Theta^2(U(g_t X), \bar{g}_t \theta) - \rho_\Theta^2(U(X), \theta))|^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \forall \theta \in \Theta\}$$

y la proposición probada. □

Proposición 5.2.2. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y consideremos una familia de distribuciones de probabilidad en (Ω, \mathcal{A}) $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ absolutamente continuas respecto a μ .

Sea $\mathcal{G}_T = \{g_t : \Omega \rightarrow \Omega \mid t \in T\}$, $T \subset \mathbb{R}$, una familia de transformaciones biyectivas y bimedibles conteniendo a la identidad y denotemos por e aquel elemento de T para el que g_e es la aplicación identidad sobre Ω .

Sea (Δ, ρ) la variedad ampliada y supongamos que ρ_Θ^2 sea una función de pérdida invariante de orden fuerte β_S .

Sea $U : \Omega \rightarrow \Theta$ un estimador de los parámetros del modelo. Entonces:

$$\beta_S \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_\theta(\tau_\theta^{\bar{g}_t \theta} \exp_{\bar{g}_t \theta}^{-1} \bar{g}_t U(X) - \exp_\theta^{-1} U(X))|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \forall \theta \in \Theta\}$$

donde $\tau_{\theta_1}^{\theta_2}$ es el transporte paralelo a lo largo de la geodésica minimal que une θ_1 y $\theta_2 \in \Theta$.

Si además U es un estimador equivariante de orden fuerte γ_S :

$$\beta_S \wedge \gamma_S \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{|E_\theta(\tau_\theta^{\bar{g}_t \theta} \exp_{\bar{g}_t \theta}^{-1} U(g_t X) - \exp_\theta^{-1} U(X))|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \forall \theta \in \Theta\}$$

Demostración. Sea $\theta \in \Theta$ arbitrario pero fijo. Se verifica que:

$$\|E_\theta(\tau_\theta^{\bar{g}_t \theta} \exp_{\bar{g}_t \theta}^{-1} \bar{g}_t U(X) - \exp_\theta^{-1} U(X))\| \leq E_\theta(\|\tau_\theta^{\bar{g}_t \theta} \exp_{\bar{g}_t \theta}^{-1} \bar{g}_t U(X) - \exp_\theta^{-1} U(X)\|)$$

Por otra parte, si fijamos $w \in \Omega$ y $\gamma > 0$, la aplicación:

$$T \ni t \mapsto \frac{\tau_\theta^{\bar{g}_t \theta} \exp_{\bar{g}_t \theta}^{-1} U(g_t X) - \exp_\theta^{-1} U(X)}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}}$$

es continua. Recordando que hemos impuesto que T sea un compacto de \mathbb{R} que contiene a e (esto no supone ninguna restricción a nuestro problema, pues

únicamente estamos interesados en perturbaciones cercanas a la identidad), podemos asegurar la existencia de $t_0 = t_0(w)$ tal que:

$$\frac{\tau_{\theta}^{\bar{g}_{t_0}\theta} \exp_{\bar{g}_{t_0}\theta}^{-1} U(g_{t_0}X) - \exp_{\theta}^{-1} U(X)}{\rho^2(g_{t_0}^*\theta, \theta)^{\frac{\gamma}{2}}} = \sup_{t \in T} \frac{\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} U(g_tX) - \exp_{\theta}^{-1} U(X)}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}}$$

Sea Z la variable aleatoria definida según:

$$Z = \frac{\tau_{\theta}^{\bar{g}_{t_0}\theta} \exp_{\bar{g}_{t_0}\theta}^{-1} U(g_{t_0}X) - \exp_{\theta}^{-1} U(X)}{\rho^2(g_{t_0}^*\theta, \theta)^{\frac{\gamma}{2}}}$$

La variable Z es integrable y domina a $\frac{\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} U(g_tX) - \exp_{\theta}^{-1} U(X)}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}}$, $\forall t \in T$. Por lo tanto, tiene sentido aplicar el Teorema de Convergencia Dominada de Lebesgue y afirmar que:

$$\begin{aligned} & \lim_{t \rightarrow e} \frac{\|E_{\theta}(\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\theta}^{-1} U(X))\|}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}} \leq \\ & \leq \lim_{t \rightarrow e} \frac{E_{\theta}(\|\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\theta}^{-1} U(X)\|)}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}} = \\ & = E_{\theta}\left(\lim_{t \rightarrow e} \frac{\|\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\theta}^{-1} U(X)\|}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}}\right) \end{aligned}$$

Pero si $\gamma < \beta_S$ el límite anterior es cero pues:

$$\lim_{t \rightarrow e} \frac{\|\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\theta}^{-1} U(X)\|}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s.}, \forall \gamma < \beta_S$$

al ser la función de pérdida invariante de orden fuerte β_S . Como $\theta \in \Theta$ arbitrario pero fijo, queda demostrado que:

$$\beta_S \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{\|E_{\theta}(\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\theta}^{-1} U(X))\|}{\rho^2(g_t^*\theta, \theta)^{\frac{\gamma}{2}}} = 0, \forall \theta \in \Theta\}$$

tal y como queríamos ver. Supongamos ahora que U sea equivariante de orden fuerte γ_S y volvamos a fijar $\theta \in \Theta$. Observemos que:

$$\|E_{\theta}(\tau_{\theta}^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} U(g_tX) - \exp_{\theta}^{-1} U(X))\| \leq$$

$$\leq \|E_\theta\left(\tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} U(g_t X) - \tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X)\right)\| + \\ + \|E_\theta\left(\tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_\theta^{-1} U(X)\right)\|$$

Hemos demostrado anteriormente que:

$$\lim_{t \rightarrow e} \frac{\|E_\theta\left(\tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_\theta^{-1} U(X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0$$

siempre que $\gamma < \beta_S$. Por lo tanto, nuestro objetivo ahora es probar que $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\|E_\theta\left(\tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \tau_\theta^{\bar{g}_t\theta} \exp_{\bar{g}_t\theta}^{-1} U(g_t X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0$$

O equivalentemente:

$$\lim_{t \rightarrow e} \frac{\|E_\theta\left(\exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\bar{g}_t\theta}^{-1} U(g_t X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0$$

Recordemos que $K = \{\bar{g}_t \theta \mid \theta \in \Theta\}$ es un conjunto compacto, tal y como hemos comentado en la proposición 5.2.1. De esa manera, debe existir $\theta' \in K$ tal que:

$$\frac{\|E_\theta\left(\exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\bar{g}_t\theta}^{-1} U(g_t X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = \\ = \sup_{\tilde{\theta} \in K} \frac{\|E_\theta\left(\exp_{\tilde{\theta}}^{-1} \bar{g}_t U(X) - \exp_{\tilde{\theta}}^{-1} U(g_t X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0$$

Por otra parte, y de igual manera a como lo hemos hecho anteriormente, $\forall t \in T$ podemos dominar:

$$\frac{\|E_\theta\left(\exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\bar{g}_t\theta}^{-1} U(g_t X)\right)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}}$$

por una variable aleatoria integrable. De esta manera, va a tener sentido aplicar el Teorema de Convergencia Dominada de Lebesgue y afirmar que $\forall \gamma < \gamma_S$:

$$\begin{aligned}
0 &\leq \lim_{t \rightarrow e} \frac{\|E_\theta(\exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\bar{g}_t\theta}^{-1} U(g_t X))\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} \leq \\
&\leq \lim_{t \rightarrow e} \frac{E_\theta(\|\exp_{\bar{g}_t\theta}^{-1} \bar{g}_t U(X) - \exp_{\bar{g}_t\theta}^{-1} U(g_t X)\|)}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} \leq \\
&\leq \lim_{t \rightarrow e} E_\theta\left(\frac{\|\exp_{\theta'}^{-1} \bar{g}_t U(X) - \exp_{\theta'}^{-1} U(g_t X)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}}\right) = \\
&= E_\theta\left(\lim_{t \rightarrow e} \frac{\|\exp_{\theta'}^{-1} \bar{g}_t U(X) - \exp_{\theta'}^{-1} U(g_t X)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}}\right) = 0
\end{aligned}$$

pues la función de pérdida es invariante de orden fuerte β_S y por tanto:

$$\lim_{t \rightarrow e} \frac{\|\exp_{\theta'}^{-1} \bar{g}_t U(X) - \exp_{\theta'}^{-1} U(g_t X)\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s., } \forall \theta', \theta \in \Theta$$

De esta manera, hemos probado que:

$$\beta_S \wedge \gamma_S \leq \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{\|E_\theta(\exp_{\bar{g}_t\theta}^{-1} U(g_t X) - \exp_{\bar{g}_t\theta}^{-1} U(X))\|}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \forall \theta \in \Theta\}$$

y finalizado la demostración de esta proposición.

□

5.3 Ejemplos

Continuando con los ejemplos tratados en los capítulos anteriores referentes a las familias $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ y $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$, en esta sección vamos a calcular los órdenes de equivariancia del estimador máximo verosímil (M.L.) de los parámetros del modelo. Podríamos haber seleccionado otros estimadores pero la conveniencia de utilizar el estimador M.L. en los procedimientos estadísticos habituales lo convierten en un ejemplo destacado a analizar.

Ejemplo 5.3.1. Volviendo al ejemplo 3.4.2, consideremos la familia de distribuciones de probabilidad $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$, el grupo multiplicativo de los reales positivos y la distancia de Rao al cuadrado como discrepancia en $\Delta = \mathbb{R} \times \mathbb{R}_+$ y como función de pérdida en $\mathbb{R}_+ \times \mathbb{R}_+$.

Sea x_1, \dots, x_k una muestra de tamaño k de una variable aleatoria X con distribución $N(\mu, 1)$ y denotemos $X = (x_1, \dots, x_k)$. Recordemos que, para esta familia, la función de verosimilitud es:

$$L(x_1, \dots, x_k; \mu) = \frac{1}{(2\pi)^{\frac{k}{2}}} \exp^{-\frac{1}{2} \sum_{i=1}^k (x_i - \mu)^2}$$

Una vez efectuados los cálculos correspondientes, obtendremos que el estimador M.L. para μ (que denotaremos $\hat{\mu}(X)$, o simplemente $\hat{\mu}$) es:

$$\hat{\mu}(X) = \frac{1}{k} \sum_{i=1}^k x_i = \bar{x}$$

donde \bar{x} indica la media muestral. Vamos a demostrar que para la familia paramétrica $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ y el grupo multiplicativo real, $\hat{\mu} = \bar{x}$ es un estimador con orden de equivariancia fuerte $+\infty$. Una vez demostrada esta afirmación, y en virtud de la proposición 5.1.1 podremos afirmar que $\hat{\mu}$ también es un estimador de orden débil $+\infty$.

Recordemos que si:

$$\begin{aligned} g_a : \mathbb{R} &\longrightarrow \mathbb{R} \quad , \quad a > 0 \\ x &\longmapsto ax \end{aligned}$$

es la acción del grupo multiplicativo de los reales positivos sobre \mathbb{R} , habíamos obtenido (ver la sección de ejemplos del capítulo 3 para más detalles) que $\overline{g_a}\mu = a\mu$, $\mu \in \mathbb{R}$. De esta manera:

$$\begin{aligned} \overline{g_a} \hat{\mu}(X) &= a \hat{\mu}(X) = \\ &= a \frac{1}{k} \sum_{i=1}^k x_i = \frac{1}{k} \sum_{i=1}^k ax_i = \hat{\mu}(g_a X) \end{aligned}$$

donde $g_a X = (g_a x_1, \dots, g_a x_k)$.

Por lo tanto, $\forall \mu' \in \mathbb{R}$:

$$\exp_{\mu'}^{-1} \overline{g_a} \hat{\mu}(X) = \exp_{\mu'}^{-1} \hat{\mu}(g_a X))$$

condición que implica que $\hat{\mu}$ es un estimador de orden fuerte $+\infty$ tal y como se quería demostrar.

Ejemplo 5.3.2. Para la familia de distribuciones de probabilidad $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$ y el grupo aditivo de los reales $\mathcal{G} = \{g_a \mid a \in \mathbb{R}\}$, (donde recordemos que la acción estaba definida según $g_a x = a + x$, $x \in \mathbb{R}$) habíamos demostrado que si X era una variable aleatoria con ley $N(0, \sigma)$ y $a \in \mathbb{R}$ era un valor próximo a 0, el elemento de $\{N(0, \sigma) \mid \sigma \in \mathbb{R}_+\}$ más próximo, según la distancia de Rao al cuadrado, a la ley de $g_a X$ es la distribución $N(0, (\frac{a^2}{2} + \sigma)^{\frac{1}{2}})$.

Sea X una muestra de tamaño k de una variable aleatoria X con distribución $N(0, \sigma)$. Recordemos que la función de verosimilitud para este modelo es:

$$L(x_1, \dots, x_k; \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp^{-\frac{1}{2\sigma^2} \sum_{i=1}^k x_i^2}$$

Sea $X = (x_1, \dots, x_k)$. Si buscamos el estimador M.L. para σ , que denotaremos $\hat{\sigma}(X)$, obtenemos que:

$$\hat{\sigma}(X) = \left(\frac{1}{k} \sum_{i=1}^k x_i^2 \right)^{\frac{1}{2}}$$

Podemos calcular ahora el orden de equivariancia del estimador $\hat{\sigma}$. Siguiendo con las notaciones anteriores, si $a \in \mathbb{R}$ sea $g_a X = (g_a x_1, \dots, g_a x_k)$. Entonces:

$$\hat{\sigma}(g_a X) = \left(\frac{1}{k} \sum_{i=1}^k (x_i + a)^2 \right)^{\frac{1}{2}} = \left(\hat{\sigma}^2(X) + 2a\bar{x} + a^2 \right)^{\frac{1}{2}}$$

donde $\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i$ es la media muestral. Por otra parte:

$$\overline{g_a} \hat{\sigma}(X) = \left(\frac{a^2}{2} + \hat{\sigma}^2(X) \right)^{\frac{1}{2}}$$

pues para valores del parámetro "a" próximos a 0 podemos definir una aplicación en el espacio de parámetros \mathbb{R}_+ dada por:

$$\sigma \mapsto \left(\frac{a^2}{2} + \sigma^2 \right)^{\frac{1}{2}}, \quad \sigma \in \mathbb{R}_+$$

Vamos a obtener el orden de equivariancia débil de $\hat{\sigma}$. Consultando Burbea [9], obtenemos que para la distribución normal univariante con media fija (en nuestro caso cero) y varianzas σ_1^2 y $\sigma_2^2 \in \mathbb{R}_+$ la distancia de Rao entre $N(0, \sigma_1)$ y $N(0, \sigma_2)$ es:

$$\rho^2((0, \sigma_1), (0, \sigma_2)) = 2 \log^2 \left(\frac{\sigma_1}{\sigma_2} \right)$$

De esta manera, $\forall \sigma' \in \mathbb{R}_+$:

$$\begin{aligned} & (\rho^2(\hat{\sigma}(g_a X), \sigma') - \rho^2(\overline{g_a} \hat{\sigma}(X), \sigma'))^2 = \\ & = \left(\log\left(\frac{\hat{\sigma}(g_a X)}{\sigma'}\right) + \log\left(\frac{\overline{g_a} \hat{\sigma}(X)}{\sigma'}\right) \right)^2 \cdot \rho^2((0, \hat{\sigma}(g_a X)), (0, \overline{g_a} \hat{\sigma}(X))) \end{aligned}$$

Como:

$$\lim_{a \rightarrow 1} \left(\log\left(\frac{\hat{\sigma}(g_a X)}{\sigma'}\right) + \log\left(\frac{\overline{g_a} \hat{\sigma}(X)}{\sigma'}\right) \right)^2 \neq 0, \text{ c.s.}, \forall \sigma' \in \mathbb{R}_+$$

únicamente es necesario que analicemos la parte correspondiente a:

$$\rho^2((0, \hat{\sigma}(g_a X)), (0, \overline{g_a} \hat{\sigma}(X)))$$

Observemos que:

$$\begin{aligned} \rho^2((0, \hat{\sigma}(g_a X)), (0, \overline{g_a} \hat{\sigma}(X))) &= \frac{1}{2} \log^2 \left(\frac{\hat{\sigma}^2(X) + 2a\bar{x} + a^2}{\frac{a^2}{2} + \hat{\sigma}^2(X)} \right) = \\ &= \frac{1}{2} \log^2 \left(1 + \frac{\frac{a^2}{2} + 2a\bar{x}}{\frac{a^2}{2} + \hat{\sigma}^2(X)} \right) \end{aligned}$$

Por ser a un valor próximo a 0, podemos afirmar que:

$$\rho^2((0, \hat{\sigma}(g_a X)), (0, \overline{g_a} \hat{\sigma}(X))) = \frac{1}{2} \left(\sum_{j=1}^{\infty} (-1)^{j+1} \left(\frac{\frac{a^2}{2} + 2a\bar{x}}{\frac{a^2}{2} + \hat{\sigma}^2(X)} \right)^j \right)^2, \text{ c.s.}$$

Pero:

$$\lim_{a \rightarrow 0} \frac{\frac{\frac{a^2}{2} + 2a\bar{x}}{\frac{a^2}{2} + \hat{\sigma}^2(X)}}{a} = \frac{2\bar{x}}{\hat{\sigma}^2(X)} \neq 0, \text{ c.s.}$$

con lo que $\hat{\sigma}$ es un estimador equivariante de orden débil 1.

Ejemplo 5.3.3. Continuando con el ejemplo anterior, consideremos el estimador de σ dado por:

$$\tilde{\sigma}(x_1, \dots, x_k) = \left(\frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

Para $\tilde{\sigma}$ se verifica:

$$\begin{aligned}
\tilde{\sigma}(g_a X) &= \left(\frac{1}{k-1} \sum_{i=1}^k \left((x_i + a) - \frac{1}{k} \sum_{j=1}^k (x_j + a) \right)^2 \right)^{\frac{1}{2}} = \\
&= \left(\frac{1}{k-1} \sum_{i=1}^k \left(x_i + a - \left(\frac{1}{k} \sum_{j=1}^k x_j \right) - a \right)^2 \right)^{\frac{1}{2}} = \tilde{\sigma}(X)
\end{aligned}$$

Por otra parte, el desplazamiento sobre el espacio de parámetros del estimador $\tilde{\sigma}$ es:

$$\overline{g_a} \tilde{\sigma}(X) = \left(\frac{a^2}{2} + \tilde{\sigma}^2(X) \right)^{\frac{1}{2}}$$

Sea ahora $\sigma' \in \mathbb{R}_+$. De igual manera a como ocurría en el ejemplo anterior, se verifica que:

$$\begin{aligned}
&(\rho^2(\tilde{\sigma}(g_a X), \sigma') - \rho^2(\overline{g_a} \tilde{\sigma}(X), \sigma'))^2 = \\
&= \left(\log\left(\frac{\tilde{\sigma}(g_a X)}{\sigma'}\right) + \log\left(\frac{\overline{g_a} \tilde{\sigma}(X)}{\sigma'}\right) \right)^2 \cdot \rho^2((0, \tilde{\sigma}(g_a X)), (0, \overline{g_a} \tilde{\sigma}(X)))
\end{aligned}$$

Nuevamente:

$$\lim_{a \rightarrow 1} \left(\log\left(\frac{\hat{\sigma}(g_a X)}{\sigma'}\right) + \log\left(\frac{\overline{g_a} \hat{\sigma}(X)}{\sigma'}\right) \right)^2 \neq 0, \text{ c.s., } \forall \sigma' \in \mathbb{R}_+$$

con lo que únicamente analizaremos la parte correspondiente a:

$$\rho^2((0, \tilde{\sigma}(g_a X)), (0, \overline{g_a} \tilde{\sigma}(X)))$$

Se verifica que:

$$\begin{aligned}
&\rho^2((0, \tilde{\sigma}(g_a X)), (0, \overline{g_a} \tilde{\sigma}(X))) = \\
&= 2 \log^2 \left(\frac{\tilde{\sigma}(g_a X)}{\overline{g_a} \tilde{\sigma}(X)} \right) = \frac{1}{2} \log^2 \left(1 + \frac{\frac{a^2}{2}}{\tilde{\sigma}(X)^2} \right) \\
&= \frac{1}{2} \left(\sum_{j=1}^{\infty} (-1)^{j+1} \left(\frac{\frac{a^2}{2}}{\tilde{\sigma}(X)^2} \right)^j \right)^2, \text{ c.s.}
\end{aligned}$$

Como:

$$\lim_{a \rightarrow 0} \frac{\frac{\frac{a^2}{2}}{\tilde{\sigma}(X)^2}}{a^2} = \frac{1}{2\tilde{\sigma}(X)^2} \neq 0, \text{ c.s.}$$

podemos afirmar que $\tilde{\sigma}$ es un estimador de σ equivariante de orden débil 2.

Con este ejemplo finalizamos el capítulo 5 dedicado al estudio de los estimadores. Queda así completado nuestro objetivo de reproducir, en la medida de lo posible, las definiciones básicas de invariancia clásica. En el siguiente capítulo veremos como además de haber recuperado estos conceptos, los podemos relacionar entre ellos de igual manera a como lo hizo Lehmann [27] en sus resultados de 1951 .

Capítulo 6

Lehmann y la invariancia en orden

Con el capítulo anterior hemos dado por finalizado el objetivo principal de este trabajo: reproducir en orden, y en la medida de lo posible, las ideas de la teoría de invariancia clásica. Los estimadores de los parámetros de un modelo han sido el último elemento a analizar, pero antes hemos definido y estudiado lo que entendemos por una familia de distribuciones de probabilidad invariante en orden y también por una función de pérdida invariante en orden débil y fuerte.

En este capítulo nos proponemos utilizar todos estos conceptos para volver a relacionar la invariancia en orden, el riesgo y el sesgo pero ahora desde un punto de vista diferente, esto es, basándonos en el trabajo de Lehmann [27] de 1951.

6.1 Estimadores insesgados en sentido clásico

Sea $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en un espacio medible (Ω, \mathcal{A}) . En los términos de Lehmann, sea U una decisión y $\delta(U(X), \theta)$ una pérdida asociada U cuando el verdadero valor del parámetro es $\theta \in \Theta$.

Definición 6.1.1. Diremos que U es *insesgada* si para cada $\theta \in \Theta$:

$$E_\theta(\delta(U(X), \vartheta)) \text{ es mínima cuando } \vartheta = \theta$$

donde en la expresión anterior E_θ indica la esperanza respecto a la distribución $P_\theta \in \mathcal{P}_\Theta$.

Recordemos que más detalles sobre decisiones y espacios de decisiones pueden ser consultados en la sección 4.1 de este trabajo.

Con esta definición en mente, y en el marco de la teoría de invariancia podemos enunciar el siguiente resultado al respecto de los procedimientos de decisión insesgados.

Proposición 6.1.1. *Sea \mathcal{G} un grupo de transformaciones medibles y biyectivas de Ω en Ω y supongamos que la familia \mathcal{P}_Θ sea invariante respecto a \mathcal{G} . Sea $\bar{\mathcal{G}}$ el grupo inducido en el espacio de parámetros e imponamos la existencia de un grupo \mathcal{G}^* de transformaciones biyectivas en el espacio de las decisiones, \mathcal{D} , homeomorfo a $\bar{\mathcal{G}}$.*

Supongamos además que la función de pérdida δ sea invariante respecto a \mathcal{G} y que éste verifique:

1. *La acción de \mathcal{G} es transitiva, esto es, dados $\theta_1, \theta_2 \in \Theta$ existe $g \in \mathcal{G}$ tal que $\theta_2 = \bar{g} \theta_1$.*
2. *\mathcal{G}^* es conmutativo.*

Entonces, si entre todos los procedimientos de decisión existe uno que minimice el riesgo uniformemente, éste es insesgado.

Este resultado caracteriza, bajo ciertas condiciones, los procedimientos de decisión insesgados. En el caso de la estimación puntual, que va a ser del que básicamente nos ocuparemos, esta caracterización nos ayudará a la hora de seleccionar un estimador u otro de los parámetros del modelo.

Sin embargo, las condiciones de la proposición 6.1.1 no tienen porque verificarse en nuestra situación de invariancia en orden. En primer lugar, debemos recordar que la familia de transformaciones el espacio muestral que nosotros consideramos no es un grupo y por lo tanto no tiene definida una operación interna. De esta manera, no tiene sentido exigir la propiedad de conmutatividad a la correspondiente familia \mathcal{G}^* . Sin embargo, podremos suavizar una vez más nuestros requerimientos tal y como se demostrará la proposición 6.2.3.

6.2 Orden de conmutatividad

En primer lugar, reestablezcamos el marco donde se van a desarrollar nuestras afirmaciones.

Sea \mathcal{P}_Θ una familia paramétrica de distribuciones de probabilidad en un espacio medible (Ω, \mathcal{A}) y absolutamente continuas respecto a μ , una medida de probabilidad en (Ω, \mathcal{A}) .

Sea $T \subset \mathbb{R}$ y $\mathcal{G}_T = \{g_t \mid t \in T\}$ una familia de transformaciones biyectivas y bimedibles de Ω en Ω conteniendo a la identidad. Sea e aquel elemento de T de manera que g_e sea la identidad sobre Ω . Sobre las transformaciones \mathcal{G}_T exigiremos que verifiquen la condición (4.1), detallada en el capítulo 4, y que tiendan a la identidad cuando $t \rightarrow e$. (Recordemos que esta última condición requería de Θ que tuviera, como mínimo, estructura de espacio topológico). Supondremos que \mathcal{P}_Θ es invariante de orden α_I respecto a \mathcal{G}_T y que Δ , la variedad ampliada, tiene estructura de variedad Riemanniana donde la distancia es la inducida por la información de Fisher. Denotémosla por ρ y sea ρ_Θ la distancia en Θ inducida por ρ .

Obviaremos la existencia de \mathcal{G}^* , el grupo homeomorfo al inducido que actúa en el espacio de decisiones, pues tal y como veremos más adelante, basta que nos preocupemos por el *orden de conmutatividad* de los elementos \bar{g}_t , $t \in T$. Debemos definir lo que entendemos por este concepto.

Definición 6.2.1. Con las notaciones precedentes, diremos que

$$\eta_W = \sup\{\eta \geq 0 \mid \lim_{t \rightarrow e} \frac{\rho_\Theta^2(\bar{g}_t \bar{g}_s \theta, \bar{g}_s \bar{g}_t \theta)}{\rho^2(g_t^* \theta, \theta)^\eta} = 0, \forall \theta \in \Theta, s \in T\}$$

es el *orden de conmutatividad débil* de \mathcal{G}_T y que:

$$\eta_S = \sup\{\eta \geq 0 \mid \lim_{t \rightarrow e} \frac{\exp_\theta^{-1} \bar{g}_t \bar{g}_s \theta - \exp_\theta^{-1} \bar{g}_s \bar{g}_t \theta}{\rho^2(g_t^* \theta, \theta)^{\frac{\eta}{2}}} = 0, \forall \theta \in \Theta, s \in T\}$$

es el *orden de conmutatividad fuerte* de \mathcal{G}_T .

Antes de pasar a demostrar la proposición 6.2.3 sobre órdenes de conmutatividad, debemos enunciar una serie de resultados previos. Aunque estos lemas corresponden al ámbito de la Geometría Diferencial, se ha optado por incluirlos en este punto pues únicamente aquí los vamos a necesitar.

Lema 6.2.1. Sea Δ una variedad Riemanniana modelada sobre un espacio de Hilbert e indiquemos por ρ la distancia Riemanniana en Δ . Sea $p \in \Delta$ y x un sistema de coordenadas en p . Entonces, si $q \rightarrow p$:

$$\rho^2(p, q) = \|x(p) - x(q)\|^2 + O(\rho^3(p, q))$$

Demostración. Sean p y $q \in \Delta$. Es conocido (ver Lang [26], entre otros) que, dados dos puntos suficientemente próximos en una variedad Riemanniana, localmente siempre existe una geodésica minimal de forma que la distancia Riemanniana entre los extremos coincide con la longitud de arco. Así, sean p y q están en la situación anterior, x un sistema de coordenadas

alrededor de $p \in \Delta$ y σ tal geodésica minimal uniendo p y q y parametrizada por la longitud de arco. Siguiendo con nuestras notaciones, vamos a considerar $t \in T$, donde T es el intervalo de \mathbb{R} que contiene a e . Sea $t > e$ de manera que $\sigma(e) = p$ y $\sigma(t) = q$. Podemos escribir:

$$x(\sigma(s)) = x(p) + v(s - e) + O((s - e)^2), \quad \forall e < s \leq t$$

donde v es el vector tangente a σ en el origen respecto al sistema de coordenadas x . De esta manera:

$$x(q) - x(p) = v(t - e) + O((t - e)^2)$$

con lo que:

$$\|x(q) - x(p)\|^2 = \|v\|^2 (t - e)^2 + O((t - e)^3)$$

Pero $\|v\| = 1$ y $t - e = \rho(p, q)$ si q está suficientemente cerca de p (pues la geodésica σ estaba parametrizada por la longitud de arco). Por lo tanto, podemos afirmar que:

$$\rho^2(p, q) = \|x(p) - x(q)\|^2 + O(\rho^3(p, q))$$

tal y como queríamos demostrar. □

Lema 6.2.2. *Sea Δ una variedad Riemanniana modelada sobre un espacio de Hilbert y denotemos por ρ la distancia Riemanniana en Δ . Sea $p \in \Delta$ y σ una curva de clase C^r , $r \geq 2$, parametrizada por la longitud de arco y de manera que $\sigma(0) = p$. Entonces:*

$$\lim_{t \rightarrow 0} \frac{\rho^2(\sigma(t), p)}{(t - 0)^2} = 1$$

Demostración. Observemos, en primer lugar, que el campo tangente a la curva, es decir $\sigma_*(\frac{d}{dt})$ es unitario al estar σ parametrizada por la longitud de arco.

Por otra parte, sea x un sistema de coordenadas en p y escojamos t suficientemente próximo a e de manera que, utilizando el lema anterior, podamos afirmar que:

$$\rho^2(p, \sigma(t)) = \|x(p) - x(\sigma(t))\|^2 + O(\rho^3(p, \sigma(t)))$$

Desarrollando en serie de Taylor y al estar t próximo a e :

$$x(\sigma(t)) = x(p) + v(t - e) + O((t - e)^2)$$

donde v es el vector tangente en el sistema de coordenadas x a la curva σ en el origen. De esta manera, podemos escribir:

$$\rho^2(p, \sigma(t)) = \|v\|^2(t - e)^2 + O((t - e)^3)$$

ya que $0 \leq \rho(p, \sigma(t)) \leq t - e$. Recordando que $\|v\| = 1$ tenemos que:

$$\rho^2(p, \sigma(t)) = (t - e)^2 + O((t - e)^3)$$

de donde:

$$\lim_{t \rightarrow 0} \frac{\rho^2(p, \sigma(t))}{(t - e)^2} = 1$$

□

Con estos resultados en mente, podemos pasar a la siguiente proposición sobre órdenes de conmutatividad.

Proposición 6.2.3. η_W y η_S son mayores o iguales que 1.

Demostración. Sea $\theta \in \Theta$ arbitrario pero fijo. Sabemos, pues así lo demostramos en el capítulo 2, que para t próximos a e la aplicación $t \mapsto \overline{g_t}\theta$ es diferenciable. De esta manera, si $s \in T$ es arbitrario pero fijo (y próximo a e , pues en caso contrario los elementos $\overline{g_s}\theta$ y $\overline{g_s}\overline{g_t}\theta$ no tienen sentido) la aplicación:

$$t \mapsto \rho_{\Theta}^2(\overline{g_t}\overline{g_s}\theta, \overline{g_s}\overline{g_t}\theta)$$

también es diferenciable al ser composición de aplicaciones diferenciables. Consideremos el desarrollo de Taylor en el origen de la aplicación anterior. Se verifica que:

$$\rho_{\Theta}^2(\overline{g_t}\overline{g_s}\theta, \overline{g_s}\overline{g_t}\theta) = O((t - e)^2)$$

pues g_e es la identidad sobre Ω y por lo tanto en $t = e$ se alcanza un mínimo local. De esta manera, podemos afirmar que

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t}\overline{g_s}\theta, \overline{g_s}\overline{g_t}\theta)}{(t - e)^{\eta}} = 0, \text{ si } \eta < 2$$

Por otra parte, recordemos que en el lema 6.2.2 hemos demostrado que siempre que la curva $t \mapsto g_t^* \theta$ está parametrizada por la longitud de arco (condición que siempre podemos suponer):

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \theta, \theta)}{(t - e)^2} = 1$$

Así:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} \overline{g_s} \theta, \overline{g_s} \overline{g_t} \theta)}{\rho^2(g_t^* \theta, \theta)^\eta} = 0, \text{ si } \eta < 1$$

Esta conclusión es independiente del punto $\theta \in \Theta$ escogido y del valor de $s \in T$, por lo tanto podemos afirmar que $\eta_W \geq 1$ tal y como queríamos ver. Observemos que podemos repetir todo este razonamiento con $\exp_{\theta}^{-1} \overline{g_t} \overline{g_s} \theta - \exp_{\theta}^{-1} \overline{g_s} \overline{g_t} \theta$ en el papel de $\rho_{\Theta}^2(\overline{g_t} \overline{g_s} \theta, \overline{g_s} \overline{g_t} \theta)$ (con la única precaución, de considerar un sistema de coordenadas x alrededor del punto $\theta \in \Theta$ y realizar los razonamientos sobre desarrollos de Taylor en el espacio de Hilbert correspondiente sobre el que está modelado Θ). De esta manera, también se verificaría que $\eta_S \geq 1$ y la proposición quedaría demostrada. \square

6.3 Estimadores insesgados en una dirección

Con el resultado anterior hemos demostrado que pese a que una familia de transformaciones no necesariamente verifique la propiedad conmutativa, por lo menos en orden 1 si que lo hace. Veremos que la conmutatividad en orden es suficiente para reproducir la proposición 6.1.1, eso si, en orden. Vamos a dividir la demostración en dos pasos, que se corresponden con las proposiciones 6.3.1 y 6.3.2.

Proposición 6.3.1. *Con las notaciones de la sección, sea \mathcal{P}_{Θ} una familia de distribuciones de probabilidad invariante de orden α_I respecto a \mathcal{G}_T . Sea η_W el orden de conmutatividad débil de \mathcal{G}_T .*

Supongamos que ρ_{Θ}^2 sea invariante débil de orden β_D y sea U un estimador de los parámetros del modelo equivariante de orden débil γ_W .

Entonces, existe $\epsilon_0 > 0$ tal que si $|s - e| < \epsilon_0$ el estimador:

$$\overline{g_s} U$$

es equivariante, en sentido débil, de orden mayor o igual que $\gamma_W \wedge \eta_W$.

Demostración. Queremos demostrar la existencia de $\epsilon_0 > 0$ de manera si $|s - e| < \epsilon_0$, el estimador $\overline{g_s} U$ es equivariante de orden débil mayor o igual que $\gamma_W \wedge \eta_W$, esto es, si $\gamma < \gamma_W \wedge \eta_W$:

$$\lim_{t \rightarrow e} \frac{\left(\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \theta') - \rho_{\Theta}^2(\overline{g_s} U(g_t X), \theta') \right)^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s., } \forall \theta', \theta \in \Theta$$

Con este objetivo, sea $s \in T$ próximo a e de manera que la aplicación $\overline{g_s}$ tenga sentido y fijemos θ' y $\theta \in \Theta$. Se cumple que:

$$\begin{aligned} & \lim_{t \rightarrow e} \frac{\left(\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \theta') - \rho_{\Theta}^2(\overline{g_s} U(g_t X), \theta') \right)^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = \\ &= \lim_{t \rightarrow e} \left\{ \frac{\left(\rho_{\Theta}(\overline{g_t} \overline{g_s} U(X), \theta') - \rho_{\Theta}(\overline{g_s} U(g_t X), \theta') \right)^2}{\rho^2(g_t^* \theta, \theta)^\gamma} \right. \\ & \quad \cdot \left(\rho_{\Theta}(\overline{g_t} \overline{g_s} U(X), \theta) + \rho_{\Theta}(\overline{g_s} U(g_t X), \theta) \right)^2 \Big\} \leq \\ & \leq \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} \cdot \left(\rho_{\Theta}(\overline{g_t} \overline{g_s} U(X), \theta) + \rho_{\Theta}(\overline{g_s} U(g_t X), \theta) \right)^2 \end{aligned}$$

donde en la última desigualdad se ha utilizado la triangular. Observemos que:

$$\lim_{t \rightarrow e} \left(\rho_{\Theta}(\overline{g_t} \overline{g_s} U(X), \theta') + \rho_{\Theta}(\overline{g_s} U(g_t X), \theta') \right)^2 = 4 \rho_{\Theta}^2(\overline{g_s} U(X), \theta) \neq 0, \text{ c.s.}$$

Por lo tanto, basta que demostremos:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s.}$$

Nuevamente, por la desigualdad triangular:

$$\begin{aligned} & \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq \\ & \leq \lim_{t \rightarrow e} \frac{\left(\rho_{\Theta}(\overline{g_t} \overline{g_s} U(X), \overline{g_s} \overline{g_t} U(X)) + \rho_{\Theta}(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X)) \right)^2}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq \\ & \leq 2 \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \overline{g_s} \overline{g_t} U(X))}{\rho^2(g_t^* \theta, \theta)^\gamma} + 2 \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} \end{aligned}$$

Pero si recordamos la definición de orden débil de conmutatividad, el primero de los dos límites anteriores es 0 *c.s.* al ser $\gamma < \eta_W$. De esta manera, debemos probar la existencia de $\epsilon_0 > 0$ tal que si $|s - e| < \epsilon_0$:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s.}$$

Debido a la continuidad de las aplicaciones $\overline{g_s}$, se verifica:

$$\lim_{s \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))} = 1, \text{ c.s.}$$

Por lo tanto, dado $\epsilon > 0$, existe $\delta > 0$ tal que:

$$\frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))} < 1 + \epsilon, \text{ c.s.}$$

siempre que $|s - e| < \delta$. Sea $M = 1 + \epsilon$ y $\epsilon_0 = \delta$. Seleccionemos s verificando $|s - e| < \epsilon_0$. Se cumple que:

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} = \\ &= \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_s} \overline{g_t} U(X), \overline{g_s} U(g_t X))}{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))} \cdot \frac{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} \leq \\ &\leq M \lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} \text{ c.s.} \end{aligned}$$

Pero este último límite es 0, *c.s.*, al ser U un estimador equivariante de orden débil γ_W . De esta manera, $\forall \gamma < \gamma_W$:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\overline{g_t} U(X), U(g_t X))}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s.}$$

Condición que implica:

$$\lim_{t \rightarrow e} \frac{\left(\rho_{\Theta}^2(\overline{g_t} \overline{g_s} U(X), \theta) - \rho_{\Theta}^2(\overline{g_s} U(g_t X), \theta) \right)^2}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0, \text{ c.s.}$$

tal y como queríamos ver. Como todo este razonamiento es independiente del punto θ escogido, queda demostrada la proposición. □

Con la proposición anterior, hemos dado el primer paso para reproducir, en orden, la proposición 6.1.1 establecida por Lehmann en 1951. Debe entenderse como un resultado previo que necesitaremos en la proposición 6.3.2. Pasamos a abordar cuestiones referentes a este resultado.

Observemos que, en los términos de la definición 6.1.1 y tomando como función de pérdida ρ_{Θ}^2 , diremos que un estimador U insesgado si y sólo si para cada $\theta \in \Theta$ la función:

$$\vartheta \mapsto E_{\theta}(\rho_{\Theta}^2(U(X), \vartheta))$$

tiene un mínimo en θ .

En términos de estimación puntual, podemos estar interesados en estimadores que continúen siendo insesgados (en el sentido Lehmann) siempre y cuando nos alejemos de θ en la dirección marcada por un cierto vector tangente $Y_{\theta} \in T_{\theta}(\Theta)$.

Definición 6.3.1. Diremos que U es un estimador *insesgado en la dirección del vector* $Y_{\theta} \in T_{\theta}(\Theta)$ si y sólo si la derivada direccional, evaluada en θ , de la aplicación:

$$\vartheta \mapsto E_{\theta}(\rho_{\Theta}^2(U(X), \vartheta))$$

en la dirección del vector Y_{θ} es cero.

Examinemos con más detalle la condición anterior. Observemos que la derivada direccional a la que se hace referencia en la definición anterior es igual a:

$$E_{\theta}(Y_{\theta} \rho_{\Theta}^2(U(X), \theta))$$

Observemos que la expresión anterior puede ser escrita en términos del gradiente de $\vartheta \mapsto \rho_{\Theta}^2(U(X), \vartheta)$, esto es:

$$E_{\theta}(Y_{\theta} \rho_{\Theta}^2(U(X), \theta)) = \langle E_{\theta}(\text{grad} \rho_{\Theta}^2(\cdot, U(X))|_{\theta}), Y_{\theta} \rangle$$

Tal y como podemos consultarse en Karcher [24] o bien las páginas 336–338 de Chavel [12]:

$$\text{grad} \rho_{\Theta}^2(\cdot, \theta_2)|_{\theta_1} = -2 \exp_{\theta_1}^{-1} \theta_2, \quad \forall \theta_1, \theta_2 \in \Theta$$

con lo que estamos el estimador U es insesgado en la dirección de $Y_{\theta} \in T_{\theta}(\Theta)$ si se verifica:

$$\langle E_{\theta}(\exp_{\theta}^{-1} U(X)), Y_{\theta} \rangle = 0$$

En este marco se va a establecer la proposición 6.3.2 donde queda reproducido, en la medida de lo posible, el resultado 6.1.1 de Lehmann.

Proposición 6.3.2. *Con las notaciones de la sección, sea \mathcal{P}_Θ una familia de distribuciones de probabilidad en un espacio de probabilidad $(\Omega, \mathcal{A}, \mu)$ invariante de orden $\alpha_I \geq 1$ respecto a una familia de transformaciones \mathcal{G}_T y absolutamente continuas respecto μ .*

Supongamos que $\frac{d}{dt}$ y E_θ se puedan intercambiar cuando sea necesario.

Sea η_W el orden de conmutatividad débil de \mathcal{G}_T y supongamos que ρ_Θ^2 sea una función de pérdida invariante de orden débil $\beta_W > 1$.

Entonces, si entre todos los estimadores equivariantes de orden débil γ_W , existe uno que minimice el riesgo uniformemente para los estimadores de orden débil mayor o igual que $\gamma_W \wedge \eta_W$, éste es insesgado en la dirección de:

$$Y_\theta = \frac{d}{dt} \bar{g}_t \theta|_{t=e}$$

esto es:

$$\langle E_\theta \left(\exp_\theta^{-1} U(X) \right), Y_\theta \rangle = 0, \quad \forall \theta \in \Theta$$

Demostración. Sea $\theta \in \Theta$ y U un estimador equivariante de orden γ_W que minimice el riesgo uniformemente para los estimadores de orden débil mayor o igual que $\gamma_W \wedge \eta_W$. Introducimos:

$$B_\theta(t) = E_\theta \left(\rho^2(\bar{g}_t \theta, U(X)) \right)$$

Queremos demostrar que:

$$\frac{d}{dt} B_\theta(t)|_{t=e} = 0$$

de donde obtendremos:

$$\langle E_\theta \left(\exp_\theta^{-1} U(X) \right), Y_\theta \rangle = 0$$

tal y como queremos ver. Como el resultado anterior es independiente del punto θ escogido, la proposición quedará demostrada.

Así, nuestro objetivo es ahora demostrar que $\frac{d}{dt} B_\theta(t)|_{t=e} = 0$. Observemos que la función de pérdida es invariante de orden débil $\beta_W > 1$. De esta manera, $\forall \beta \leq 1$:

$$\lim_{t \rightarrow e} \frac{(\rho_\Theta^2(\bar{g}_t \theta_1, \bar{g}_t \theta_2) - \rho_\Theta^2(\theta_1, \theta_2))^2}{\min\{\rho^2(g_t^* \theta_1, \theta_1)^\beta, \rho^2(g_t^* \theta_2, \theta_2)^\beta\}} = 0$$

Haciendo $\theta_1 = U(X)$ y $\theta_2 = \theta$, la afirmación anterior implica que $\forall \beta \leq 1$:

$$\lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta))^2}{\rho^2(g_t^* \theta, \theta)^\beta} = 0, \text{ c.s.}$$

Recordemos que en la proposición 6.2.2 habíamos demostrado que si la curva $t \mapsto g_t^* \theta$ está parametrizada por la longitud de arco (condición que siempre podemos suponer), entonces:

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \theta, \theta)}{(t - e)^2} = 1$$

Por lo tanto, también debe verificarse que para $\forall \beta \leq 1$:

$$\lim_{t \rightarrow e} \frac{(\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta))^2}{(t - e)^{2\beta}} = 0, \text{ c.s.}$$

De aquí, tomando $\beta = 1$, podemos afirmar que:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta)}{(t - e)} = 0, \text{ c.s.}$$

de donde:

$$\frac{d}{dt} \left(\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta) \right) \Big|_{t=e} = 0, \text{ c.s.}$$

Debido a las hipótesis de la proposición, podemos intercambiar $\frac{d}{dt}$ y E_{θ} cuando sea necesario. Así:

$$\frac{d}{dt} E_{\theta} \left(\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta) \right) \Big|_{t=e} = 0$$

Evaluamos la derivada anterior, y obtenemos que:

$$\begin{aligned} 0 &= \frac{d}{dt} E_{\theta} \left(\rho_{\Theta}^2(\bar{g}_t U(X), \bar{g}_t \theta) - \rho_{\Theta}^2(U(X), \theta) \right) \Big|_{t=e} = \\ &= -2 \langle E_{\theta} \left(\exp_{\theta}^{-1} U(X) \right), \frac{d}{dt} \bar{g}_t \theta \Big|_{t=e} \rangle + E_{\theta} \left(-2 \langle \tau_{U(X)}^{\theta} \exp_{\theta}^{-1} U(X), \frac{d}{dt} U(g_t X) \Big|_{t=e} \rangle \right) \end{aligned}$$

De esta manera, la proposición quedará demostrada si probamos que:

$$E_{\theta} \left(-2 \langle \tau_{U(X)}^{\theta} \exp_{\theta}^{-1} U(X), \frac{d}{dt} U(g_t X) \Big|_{t=e} \rangle \right) = 0$$

Consideremos la función:

$$R_\theta(t) = E_\theta\left(\rho_\Theta^2(\bar{g}_t U(X), \theta)\right)$$

Hemos visto en la proposición 6.3.1 que si U es un estimador equivariante de orden débil γ_W y η_W es el orden de conmutatividad débil de \mathcal{G}_T , podemos afirmar que existe $\epsilon_0 > 0$ tal que si $|t - e| < \epsilon_0$, $\bar{g}_t U$ es equivariante de orden débil mayor o igual que $\gamma_W \wedge \eta_W$. Consideremos t verificando $|t - e| < \epsilon_0$. Por otra parte, las hipótesis de la proposición establecen que entre todos los estimadores equivariantes de orden débil mayor o igual que $\gamma_W \wedge \eta_W$, U es el que minimiza el riesgo uniformemente. Así, ha de ser:

$$\frac{d}{dt} R_\theta(t)|_{t=e} = 0$$

esto es:

$$E_\theta\left(-2 \langle \tau_{U(X)}^\theta \exp_\theta^{-1} U(X), \frac{d}{dt} \bar{g}_t U(X)|_{t=e} \rangle\right) = 0$$

de donde se tiene que U es un estimador insesgado en la dirección de:

$$Y_\theta = \frac{d}{dt} \bar{g}_t \theta|_{t=e}$$

tal y como queríamos ver. □

Con este resultado damos por finalizado este capítulo dedicado a relacionar invariancia en orden, riesgo y sesgo siguiendo, en la medida de lo posible, el esquema marcado por Lehmann [27] en su trabajo de 1951. En éste se demuestra también un segundo resultado donde se caracterizan los estimadores equivariantes en sentido clásico como aquellos de mínimo riesgo entre todos los insesgados. A la hora de reproducir este resultado en orden han aparecido ciertas dificultades que se pretenden solucionar en un futuro próximo y así completar el estudio realizado en este capítulo.

Capítulo 7

Estimación estable

En los capítulos anteriores hemos reproducido, en la medida de lo posible, las ideas básicas de la teoría clásica de la invariancia, recuperando incluso algunos resultados destacados al respecto.

Recordemos que hemos partido de un espacio de medida $(\Omega, \mathcal{A}, \mu)$, de una familia de distribuciones de probabilidad $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ en (Ω, \mathcal{A}) , absolutamente continuas respecto μ , y de una familia de transformaciones \mathcal{G}_T en (Ω, \mathcal{A}) biyectivas, bimedibles y conteniendo a la identidad sobre Ω . A partir de los elementos de \mathcal{G}_T , construíamos una familia de transformaciones en el espacio de parámetros Θ a la que nos hemos referido como perturbaciones en Θ o bien perturbaciones de \mathcal{P}_Θ . Éstas están claramente relacionadas con \mathcal{G}_T , esto es, una transformación en el espacio muestral provoca una perturbación en \mathcal{P}_Θ y por lo tanto también en una parte del espacio de las medidas de probabilidad en (Ω, \mathcal{A}) . Sin embargo, observemos que las transformaciones en el espacio muestral juegan aquí un papel secundario pues, en realidad, el objeto de nuestro estudio han sido los desplazamientos en \mathcal{P}_Θ .

En este capítulo vamos a ver que podríamos ir aún más allá y suponer que la familia \mathcal{P}_Θ ha sido modificada por una perturbación sin importarnos cual sea su origen aunque, eso sí, le exigiremos que verifique ciertas condiciones de regularidad.

7.1 Familias localmente 1-paramétricas

Consideremos un espacio medible (Ω, \mathcal{A}) y sea $\mathcal{M}^1(\Omega, \mathcal{A})$ el conjunto de todas las medidas de probabilidad en (Ω, \mathcal{A}) .

Sea Θ una n -variedad \mathcal{C}^∞ y supongamos que exista una aplicación localmente inyectiva:

$$\begin{aligned} P : \Theta &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}) \\ \theta &\longmapsto P_\theta \end{aligned}$$

Denotemos por \mathcal{P}_Θ la familia de distribuciones de probabilidad $\{P_\theta \mid \theta \in \Theta\}$ en el espacio medible (Ω, \mathcal{A}) . Tal y como hemos comentado, las perturbaciones de \mathcal{P}_Θ que hasta ahora hemos considerado se han obtenido vía una familia de transformaciones en el espacio muestral. Sin embargo, no únicamente la muestra podría estar sujeta a perturbaciones. El mero hecho de seleccionar una familia paramétrica concreta a la hora de modelizar unos datos también puede verse como una perturbación de la verdadera familia, aunque ahora en $\mathcal{M}^1(\Omega, \mathcal{A})$.

Nuestro objetivo es cuantificar, inspirándonos en las ideas de la sección 3.2, las perturbaciones del modelo estadístico paramétrico $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ que pudieran existir. Tal y como veremos más adelante, el origen de éstas no es relevante pero las condiciones de regularidad que se detallan en la siguiente definición sí que lo son.

Sea T un intervalo abierto de \mathbb{R} y consideremos una colección de intervalos abiertos contenidos en T , que indicaremos $I_\Theta = \{I_\theta \mid \theta \in \Theta\}$. Supondremos que $\exists e \in T$ tal que $e \in I_\theta$, $\forall \theta \in \Theta$ y que consideramos en $\mathcal{M}^1(\Omega, \mathcal{A})$ la topología dada por la variación total.

Definición 7.1.1. *Una familia localmente 1-paramétrica de perturbaciones sobre \mathcal{P}_Θ es una colección de transformaciones:*

$$\begin{aligned} I_\theta &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}) \\ t &\longmapsto v_t(P_\theta) \end{aligned}$$

de manera que:

1. $\forall \theta \in \Theta$, $v_e(P_\theta) = P_\theta$.
2. $\forall \theta \in \Theta$, $\lim_{t \rightarrow e} v_t(P_\theta) = P_\theta$.

Al respecto de la definición anterior, debemos comentar que el comportamiento de continuidad en el parámetro exigido en la condición 2 es una propiedad deseable para cualquier familia de perturbaciones. Como ejemplo, recordemos que para la situación de invariancia en orden esta condición se verifica. En general los intervalos I_θ no tienen porque coincidir, aunque pueden encontrarse ejemplos destacados en los que esto ocurra.

Ejemplo 7.1.1. En el marco de la invariancia en orden, consideremos la familia de perturbaciones en \mathcal{P}_Θ inducidas por \mathcal{G}_T . Aquí estamos suponiendo que \mathcal{G}_T verifica las condiciones habituales detalladas de los capítulos 3 y siguientes.

Si Θ es una variedad compacta, podremos recubrirla por una colección finita de conjuntos compactos y considerar el ínfimo, sobre cada uno de los compactos, de los $t > e$ para el existiría una única proyección de $g_t^*\theta$ en Θ según la discrepancia.

Sean s_1, \dots, s_N estos ínfimos. Tomando $s = \min\{t_1, \dots, t_N\}$ e $I = [e, s)$ hemos construido una familia uniparamétrica de perturbaciones sobre \mathcal{P}_Θ que viene dada por:

$$\begin{aligned} I &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}, \mu) & , \quad \forall \theta \in \Theta \\ t &\longmapsto P_{g_t^*\theta} \end{aligned}$$

donde P es la aplicación localmente inyectiva que hemos introducido al principio de esta sección.

Disponemos de otro ejemplo de transformaciones en $\mathcal{M}^1(\Omega, \mathcal{A}, \mu)$ para el que los I_θ coinciden.

Ejemplo 7.1.2. Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de probabilidad y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\} \subset \mathcal{M}^1(\Omega, \mathcal{A}, \mu)$ donde Θ , el espacio de parámetros, es una variedad \mathcal{C}^∞ .

El teorema de Radon-Nikodym informa que existen $\frac{dP_\theta}{d\mu}$, que denotaremos $p(\cdot; \theta)$. Observemos que $\forall \theta \in \Theta$, $\sqrt{p(\cdot; \theta)} \in L^2(\Omega, \mathcal{A}, \mu)$.

Sea A un operador lineal de $L^2(\Omega, \mathcal{A}, \mu)$ de manera que $A^* = -A$ (aquí A^* indica el operador adjunto de A) y $t \in \mathbb{R}$. Consideremos la familia de operadores lineales:

$$\begin{aligned} T_t : L^2(\Omega, \mathcal{A}, \mu) &\longrightarrow L^2(\Omega, \mathcal{A}, \mu) \\ f &\longmapsto \exp^{tA} f \end{aligned}$$

Observemos que T_t son operadores unitarios pues $(T_t)^* = \exp^{-tA}$ y por lo tanto $(T_t)^* \circ T_t = T_t \circ (T_t)^* = I$. De esta manera, $\forall \theta \in \Theta$, $\|T_t(\sqrt{p(\cdot; \theta)})\| = 1$ con lo que podremos definir una nueva distribución de probabilidad en (Ω, \mathcal{A}) dada por:

$$\xi_t^\theta(B) = \int_{\Omega} T_t(\sqrt{p(\omega; \theta)}) d\mu(\omega), \quad \forall B \in \mathcal{A}$$

Identifiquemos $T_t(\sqrt{p(\cdot; \theta)})$ con ξ_t^θ . Entonces, la familia de aplicaciones:

$$\begin{aligned}\mathbb{R} &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}) \quad , \quad \theta \in \Theta \\ t &\longmapsto \xi_t^\theta\end{aligned}$$

es una perturbación uniparamétrica de \mathcal{P}_Θ . Observemos que hemos omitido el término localmente pues aquí $I_\theta = \mathbb{R}$, $\forall \theta \in \Theta$.

El último ejemplo que presentamos sobre perturbaciones es realizarlas a través de un núcleo estocástico. Recordemos la definición.

Definición 7.1.2. Sean $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ dos espacios medibles. Diremos que:

$$K : \Omega_1 \times \mathcal{A}_2 \longrightarrow \mathbb{R}_+$$

es un *núcleo estocástico* de $(\Omega_1, \mathcal{A}_1)$ en $(\Omega_2, \mathcal{A}_2)$ si y sólo si:

SK 1 La aplicación $K(\cdot, A_2) : \Omega_1 \rightarrow \mathbb{R}_+$ es \mathcal{A}_1 -medible, $\forall A_2 \in \mathcal{A}_2$.

SK 2 La aplicación $K(w_1, \cdot) : \mathcal{A}_2 \rightarrow \mathbb{R}_+$ es una medida de probabilidad en $(\Omega_2, \mathcal{A}_2)$, $\forall w_1 \in \Omega_1$.

Definición 7.1.3. Si en la definición anterior $(\Omega_1, \mathcal{A}_1) = (\Omega_2, \mathcal{A}_2)$ diremos que K es un *núcleo estocástico* en $(\Omega_1, \mathcal{A}_1)$.

Ejemplo 7.1.3. Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y:

$$K : \Omega \times \mathcal{A} \longrightarrow \mathbb{R}_+$$

un núcleo estocástico en (Ω, \mathcal{A}) y P una medida de probabilidad en (Ω, \mathcal{A}) . Consideremos la aplicación:

$$\begin{aligned}KP : \mathcal{A} &\longrightarrow [0, 1] \\ B &\longmapsto \int_{\Omega} K(w, B) dP(w)\end{aligned}$$

Veamos, en primer lugar, que está bien construida. Por la definición 7.1.3 sabemos que $\forall w \in \Omega$, $K(w, \cdot)$ es una medida de probabilidad en (Ω, \mathcal{A}) con lo que:

$$0 \leq K(w, \cdot) \leq 1$$

De esta manera, $\forall B \in \mathcal{A}$:

$$0 \leq \int_{\Omega} K(w, B) dP(w) \leq \int_{\Omega} 1 dP(w) = 1$$

tal y como queríamos ver. Vamos a ver ahora que además KP es un elemento de $\mathcal{M}^1(\Omega, \mathcal{A})$, esto es una probabilidad en (Ω, \mathcal{A}) .

Observemos que $KP(\Omega) = 1$ pues:

$$KP(\Omega) = \int_{\Omega} K(w, \Omega) dP(w) = \int_{\Omega} 1 dP(w) = 1$$

según la condición SK 2. Ésta también informa que $KP(\emptyset) = 0$ pues:

$$KP(\emptyset) = \int_{\Omega} K(w, \emptyset) dP(w) = \int_{\Omega} 0 dP(w) = 0$$

al ser $K(w, \emptyset) = 0, \forall w \in \Omega$.

Por otra parte, KP es una función de conjunto aditiva pues dados $A, B \in \mathcal{A}$ tales que $A \cap B = \emptyset$, se verifica que:

$$\begin{aligned} KP(A \cup B) &= \int_{\Omega} K(w, A \cup B) dP(w) = \int_{\Omega} K(w, A) + K(w, B) dP(w) = \\ &= \int_{\Omega} K(w, A) dP(w) + \int_{\Omega} K(w, B) dP(w) = KP(A) + KP(B) \end{aligned}$$

donde se ha vuelto a utilizar SK 2. Para probar que KP es un elemento de $\mathcal{M}^1(\Omega, \mathcal{A})$, únicamente nos queda por verificar que KP es σ -aditiva.

Sea $\{A_n\}_{n \geq 1}$ una sucesión creciente de elementos de \mathcal{A} . Entonces:

$$\begin{aligned} \lim_{n \rightarrow \infty} KP(A_n) &= \lim_{n \rightarrow \infty} \int_{\Omega} K(w, A_n) dP(w) = \int_{\Omega} \lim_{n \rightarrow \infty} K(w, A_n) dP(w) = \\ &= \int_{\Omega} K(w, \cup_{n=1}^{\infty} A_n) dP(w) = KP(\cup_{n=1}^{\infty} A_n) \end{aligned}$$

según el teorema de Convergencia Monótona pues, por SK 1, $\{K(\cdot, A_n)\}_{n \geq 1}$ es una sucesión creciente de variables aleatorias integrables de límite, también integrable, $K(\cdot, \cup_{n=1}^{\infty} A_n)$.

De esta manera, hemos construido una aplicación:

$$\begin{array}{ccc} \mathcal{M}^1(\Omega, \mathcal{A}) & \longrightarrow & \mathcal{M}^1(\Omega, \mathcal{A}) \\ P & \longmapsto & KP \end{array}$$

en la que cada probabilidad P queda perturbada a través del núcleo estocástico K .

Sea ahora $P_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en (Ω, \mathcal{A}) . Sea $T \subset \mathbb{R}$ y $\{K_t\}_{t \in T}$ una familia de núcleos estocásticos en (Ω, \mathcal{A}) .

Supongamos que $\exists e \in T$ de manera que $K_e P = P$ y que $\lim_{t \rightarrow e} K_t P = P$, $\forall P \in \mathcal{M}^1(\Omega, \mathcal{A})$. Observemos que en tal situación las transformaciones:

$$\begin{aligned} T &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}) \\ t &\longmapsto K_t P_\theta \end{aligned}$$

es una familia uniparamétrica de perturbaciones sobre \mathcal{P}_Θ . Se ha omitido el término localmente porque, de nuevo, $I_\theta = T$, $\forall \theta \in \Theta$.

Ejemplo 7.1.4. Recuperemos el ejemplo anterior pero ahora considerando un núcleo estocástico concreto.

Sea (Ω, \mathcal{A}) un espacio medible y $g : \Omega \longrightarrow \Omega$ una aplicación biyectiva. Introducimos:

$$K_g(w, B) = \mathbf{1}_{g^{-1}(B)}(w), \quad w \in \Omega, B \in \mathcal{A}$$

donde $\mathbf{1}$ es la función indicador. Observemos que K_g es un núcleo estocástico en (Ω, \mathcal{A}) cumpliendo que $\forall P \in \mathcal{M}^1(\Omega, \mathcal{A})$:

$$\begin{aligned} K_g P(B) &= \int_{\Omega} K_g(w, B) dP(w) = \int_{\Omega} \mathbf{1}_{g^{-1}(B)}(w) dP(w) = \\ &= P(g^{-1}(B)) = P^g(B), \quad \forall B \in \mathcal{A} \end{aligned}$$

es decir, perturbamos una distribución de probabilidad por una transformación biyectiva obteniendo la ley inducida. De esta manera, vemos como los núcleos estocásticos permiten expresar de una manera alternativa la situación expuesta en el capítulo 3 donde partiendo de P_θ y de una transformación en el espacio muestral $g_t \in \mathcal{G}_T$ intentábamos encontrar el elemento de la familia original que estuviera a discrepancia mínima de $P_\theta^{g_t}$ o, con las notaciones de este ejemplo, de $K_{g_t} P_\theta$.

Con este último ejemplo finalizamos la exposición sobre posibles perturbaciones que pudieran darse en $\mathcal{M}^1(\Omega, \mathcal{A})$. Los capítulos precedentes han estado dedicados a aquellas que se obtenían en la situación de invariancia en orden. Pero, tal y como hemos visto en esta sección, ésta no es la única posibilidad.

7.2 Familias y pérdidas estables

En el resto de este capítulo, y con ayuda de los capítulos precedentes, pretendemos apuntar las ideas para realizar un estudio sobre las perturbaciones arbitrarias de las medidas de probabilidad en un espacio medible (Ω, \mathcal{A}) que sean absolutamente continuas respecto a una medida σ -finita de referencia μ en (Ω, \mathcal{A}) .

Sea así $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y supongamos que disponemos de una familia de distribuciones de probabilidad $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ en (Ω, \mathcal{A}) absolutamente continuas respecto μ . Tal y como hicimos en la sección anterior, supondremos que la aplicación:

$$\begin{aligned} P : \Theta &\longrightarrow \mathcal{M}^1(\Omega, \mathcal{A}, \mu) \\ \theta &\longmapsto P_\theta \end{aligned}$$

es localmente inyectiva, donde ahora $\mathcal{M}^1(\Omega, \mathcal{A}, \mu)$ es el conjunto de probabilidades en (Ω, \mathcal{A}) absolutamente continuas respecto μ . De esta manera, podremos identificar localmente los elementos de \mathcal{P}_Θ con sus etiquetas e imponer que Θ sea una subvariedad n -dimensional de una variedad Riemanniana Δ , que llamaremos ambiente. Un ejemplo de tal Δ lo tenemos en el espacio de Hilbert $L^2(\Omega, \mathcal{A}, \mu)$ identificando cada elemento de \mathcal{P}_Θ , o de Θ , con dos veces la raíz cuadrada positiva de su derivada de Radon–Nikodym respecto μ .

Sea X un campo vectorial \mathcal{C}^∞ en Δ y $\theta \in \Theta$. Observemos que las curvas integrales de X , que denotaremos v , con la condición inicial $v(e; \theta) = \theta$, $e \in \mathbb{R}$, definen una familia localmente 1-paramétrica de transformaciones en Δ que viene dada por:

$$I_\theta \ni t \longmapsto v(t; \theta) \in \Delta$$

Denotemos $v(t; \theta)$ como $v_t(\theta)$ y consideremos una discrepancia, δ , en Δ . El teorema 2.2.2 informa que para t próximos a e , existe $\theta' \in \Theta$ de manera que:

$$\delta(v_t(\theta), \theta') = \min_{\vartheta \in \Theta} \delta(v_t(\theta), \vartheta) \quad (7.1)$$

Por consonancia con las notaciones de este trabajo, denotaremos θ' como $\overline{v}_t(\theta)$. Estas consideraciones nos llevan a introducir el siguiente concepto:

Definición 7.2.1. Diremos que \mathcal{P}_Θ es *estable de orden α_E* respecto al campo X de clase \mathcal{C}^∞ si y sólo si:

$$\alpha_E = \sup\{\alpha \geq 0 \mid \lim_{t \rightarrow e} \frac{\delta(v_t(\theta), \overline{v}_t(\theta))}{\delta(v_t(\theta), \theta)^\alpha} = 0, \forall \theta \in \Theta\}$$

En este punto, observemos que una familia invariante de orden α_I respecto a una familia de transformaciones $\mathcal{G}_T = \{g_t \mid t \in T\}$ cumpliendo las condiciones habituales de los capítulos anteriores, también es estable de orden α_I respecto al campo $\frac{d}{dt}g_t^*\theta|_{t=e}$. Todo parece indicar que la invariancia en orden podría enmarcarse dentro de un concepto más amplio que podemos llamar *de estabilidad en orden*. Éste se encuentra en su fase inicial, debiéndose entender este capítulo como una breve presentación al respecto.

Observemos que de igual manera a como vimos en el capítulo 3, se verifica que el orden de estabilidad no depende de la variedad ambiente si la discrepancia escogida es una f -divergencia. Más detalles al respecto deben consultarse en la demostración de la proposición 3.3.2, que podremos adaptar a este nuevo entorno.

Recuperando las ideas del capítulo 4, también será posible definir lo que entendemos por una función de pérdida estable de órdenes β_W y β_S .

En el contexto de los modelos estadísticos paramétricos sea:

$$L : \Theta \times \Theta \longrightarrow [0, +\infty)$$

una función de pérdida.

Definición 7.2.2. Diremos que L es *estable de orden débil* β_W respecto al campo $\mathcal{C}^\infty X$ si y sólo si:

$$\beta_W = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{(L(\overline{v}_t(\theta_1), \overline{v}_t(\theta_2)) - L(\theta_1, \theta_2))^2}{\min(\delta(v_t(\theta_1), \theta_1)^\beta, \delta(v_t(\theta_2), \theta_2)^\beta)} = 0, \forall \theta_1, \theta_2 \in \Theta\}$$

Definición 7.2.3. Diremos que L es *estable de orden fuerte* β_S respecto al campo X de clase \mathcal{C}^∞ si y sólo si:

$$\begin{aligned} \beta_F = \sup\{\beta \geq 0 \mid \lim_{t \rightarrow e} \frac{\tau_{\theta_1}^{\overline{v}_t(\theta_1)} \text{grad } L(\cdot, \overline{v}_t(\theta_2))|_{\overline{v}_t(\theta_1)}}{\min(\delta(v_t(\theta_1), \theta_1)^{\frac{\beta}{2}}, \delta(v_t(\theta_2), \theta_2)^{\frac{\beta}{2}})} - \\ - \frac{\text{grad } L(\cdot, \theta_2)|_{\theta_1}}{\min(\delta(v_t(\theta_1), \theta_1)^{\frac{\beta}{2}}, \delta(v_t(\theta_2), \theta_2)^{\frac{\beta}{2}})} = 0_{P_{\theta_1}}, \forall \theta_1, \theta_2 \in \Theta\} \end{aligned}$$

siendo $0_{\theta_1} \in T_{\theta_1}(\Theta)$, $\text{grad } L(\cdot, \theta_2)|_{\theta_1}$ el gradiente de la función $L(\cdot, \theta_2)$ en el punto θ_1 y:

$$\tau_{\theta_1}^{\overline{v}_t(\theta_1)} : T_{\overline{v}_t(\theta_1)}(\Theta) \longrightarrow T_{\theta_1}(\Theta)$$

es el transporte paralelo desde $T_{\overline{v}_t(\theta_1)}(\Theta)$ a $T_{\theta_1}(\Theta)$ a lo largo de una geodésica que una $\overline{v}_t(\theta_1)$ y θ_1 .

De igual manera a como ocurría con las familias, observemos que las funciones de pérdida invariantes de órdenes β_W y β_S respecto a \mathcal{G}_T también son estables de los mismos órdenes respecto al campo $\frac{d}{dt}g_t^*\theta|_{t=e}$.

Antes de continuar debemos hacer notar que hasta el momento el campo X es arbitrario. Sin embargo, observemos que para la situación de invariancia clásica, los campos tangentes a las curvas $t \mapsto \bar{g}_t\theta$, $\theta \in \Theta$ son de *Killing*. Esto indica que como candidatos a X podríamos tomar campos de Killing. En la siguiente sección, dedicada a los estimadores estables, se pretende esbozar la conveniencia de considerar este tipo de campos. Es de esperar que en futuras investigaciones estas consideraciones queden completadas.

7.3 Estimadores estables

El objetivo de esta sección es la obtención de criterios de estabilidad de un estimador respecto a la perturbación introducida por un campo X de clase \mathcal{C}^∞ . Esperaremos que tal y como ha ocurrido con la familia y la función de pérdida, este criterio de estabilidad sea el de equivariancia en orden cuando el campo considerado es $X(\theta) = (\frac{d}{dt}g_t^*\theta)|_{t=e}$.

Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida y $\mathcal{P}_\Theta = \{P_\theta \mid \theta \in \Theta\}$ una familia paramétrica de distribuciones de probabilidad en (Ω, \mathcal{A}) absolutamente continuas respecto μ . Supongamos que el espacio de parámetros Θ es una n -subvariedad de la variedad ambiente Δ . Tal y como hemos comentado en repetidas ocasiones, la variedad Δ se considerará Riemanniana siendo la distancia Riemanniana, ρ , en Δ la inducida por la matriz de Información de Fisher.

Sea X un campo vectorial \mathcal{C}^∞ en Δ y, con las notaciones de la sección anterior, $\{v_t\}_t$ la familia localmente uniparamétrica de transformaciones en Δ determinadas por X y $\{\bar{v}_t\}_t$ las transformaciones en Θ asociadas a $\{v_t\}_t$ que han sido introducidas en (7.1). Finalmente, sean:

$$U : \Omega^k \longrightarrow \Theta$$

un estimador de los parámetros del modelo y δ una discrepancia en Δ .

La existencia de perturbaciones detallada anteriormente invita a considerar dos nuevos estimadores. Por una parte, podemos pensar que si U estima a θ , $\bar{v}_t \circ U$ debe hacer lo propio con $\bar{v}_t\theta$. Pero también podemos incorporar la información referente a que la familia ha sido perturbada, considerar que la ley es $P_{\bar{v}_t\theta}$ en lugar de P_θ y utilizar U para estimar $\bar{v}_t\theta$.

Observemos que en tal caso disponemos de dos distribuciones de probabilidad en el espacio medible $(\Theta, \mathcal{B}(\Theta))$, $P_\theta^{\bar{v}_t \circ U}$ y $P_{\bar{v}_t\theta}^U$ (donde $\mathcal{B}(\Theta)$ es la σ -álgebra de

Borel en Θ). Un posible criterio de estabilidad para el estimador T es exigir que una medida de probabilidad y la otra sean *lo más parecidas posible*. Obviamente, debemos concretar el significado de este término.

Definición 7.3.1. Diremos que el estimador U es *estable respecto a la perturbación introducida por el campo X de clase \mathcal{C}^∞* si y sólo si:

$$\delta(P_\theta^{\bar{v}_t \circ U}, P_{\bar{v}_t \theta}^U) = 0, \quad \forall t \in I_\theta$$

Si disponemos de su forma explícita para el modelo paramétrico estudiado, parece razonable seleccionar $\delta = \rho^2$. Sin embargo, también podemos tomar $\delta = d_H^2$ o bien cualquier otra f -divergencia. Dependerá de cada situación concreta una elección u otra, pero tomando esta clase de discrepancias sabemos que el orden de estabilidad no dependerá de la selección realizada. Esta afirmación de demuestra adaptando los argumentos de la proposición 3.3.2 a nuestro entorno actual.

Sin embargo, podríamos estar interesados en medir la proximidad entre medidas de probabilidad a través de otras cantidades que no fueran f -divergencias. Una posibilidad sería escoger una distancia basada en la variabilidad total (ver Rao [35] o Cuadras y Fortiana [15] para más detalles), esto es:

$$\begin{aligned} \delta(P, Q) = & \int_{\Theta} \int_{\Theta} \rho_{\Theta}^2(p, q) P(dp) Q(dq) - \\ & - \frac{1}{2} \int_{\Theta} \int_{\Theta} \rho_{\Theta}^2(p, q) P(dp) P(dq) - \frac{1}{2} \int_{\Theta} \int_{\Theta} \rho_{\Theta}^2(p, q) Q(dp) Q(dq) \end{aligned}$$

Sin embargo, debemos hacer notar que al trabajar con esta cantidad, o cualquier otra que no sea localmente (salvo una constante de proporcionalidad) la distancia de Rao cuadrática, los órdenes de estabilidad deberán ir referidos a δ .

Ejemplo 7.3.1. En la situación de invariancia clásica, $P_\theta^{\bar{v}_t \circ U}$ es la ley del estimador $\bar{g}_t \circ U$ mientras que $P_{\bar{v}_t \theta}^U$ es la ley del estimador $U \circ g_t$ suponiendo, en ambos casos, que la verdadera medida es P_θ .

Si la familia \mathcal{P}_Θ es invariante respecto a la acción de un grupo de transformaciones $\mathcal{G}_T = \{g_t \mid t \in T\}$ y el estimador U es equivariante respecto a este grupo, sabemos que $\forall t \in T$:

$$\bar{g}_t U(x) = U(g_t x), \quad \forall x \in \Omega$$

De esta manera las leyes de uno y otro estimador coincidirán, con lo que U sería un estimador estable respecto a la perturbación introducida por el campo $\theta \mapsto \frac{d}{dt} \bar{g}_t \theta|_{t=e}$, $\theta \in \Theta$.

Sin embargo, para la invariancia en orden la situación no va a ser tan favorable.

Ejemplo 7.3.2. Para la situación de invariancia en orden, $P_{\theta}^{\bar{g}_t \circ T}$ es la ley del estimador $\bar{g}_t \circ T$ donde se supone que la distribución es P_{θ} . Por otra parte, $P_{\bar{g}_t \theta}^T$ es la ley del estimador T suponiendo que la verdadera medida fuera $P_{\bar{g}_t \theta}$. En general, una y otra no tienen porque coincidir, con lo que la medida de discrepancia $P_{\theta}^{\bar{g}_t \circ T}$ y $P_{\bar{g}_t \theta}^T$ no será nula. Analicemos con algo más de detalle esta situación.

Supongamos que la familia \mathcal{P}_{Θ} sea invariante de orden α_I y sea U un estimador equivariante de orden fuerte γ_S . Supongamos, para simplificar las expresiones que $U : \Omega \rightarrow \Theta$. Tal y como hemos comentado anteriormente, debemos comparar $P_{\theta}^{\bar{g}_t \circ U}$ y $P_{\bar{g}_t \theta}^U$ a través de la discrepancia en Θ , esto es, debemos analizar:

$$\rho^2(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U)$$

conociendo que:

$$\gamma_S = \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{\exp_{\theta'}^{-1} U(g_t X) - \exp_{\theta'}^{-1} \bar{g}_t U(X)}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s. } \forall \theta', \theta \in \Theta\}$$

Siguiendo las ideas detalladas en la sección 5.1, la condición de equivariancia fuerte de un estimador implica que $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\rho_{\Theta}^2(U(g_t X), \bar{g}_t U(X))}{\rho^2(g_t^* \theta, \theta)^{\frac{\gamma}{2}}} = 0, \text{ c.s. } \forall \theta \in \Theta$$

donde recordemos que el c.s. de la definición es respecto a μ . Podemos escribir la condición anterior en términos de medidas de probabilidad. Obtendremos que $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\rho^2(P_{g_t^* \theta}^U, P_{\theta}^{\bar{g}_t \circ U})}{\rho^2(g_t^* \theta, \theta)^{\gamma}} = 0 \quad (7.2)$$

donde en la ecuación anterior se utilizan las etiquetas cuando es posible identificarlas con las probabilidades a las que representan. En caso contrario, se mantienen las medidas.

Podemos ahora comparar $P_{\theta}^{\bar{g}_t \circ U}$ y $P_{\bar{g}_t \theta}^U$ a través de ρ^2 . Observemos que por la desigualdad triangular:

$$\rho(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U) \leq \rho(P_{\theta}^{\bar{g}_t \circ U}, P_{g_t^* \theta}^U) + \rho(P_{g_t^* \theta}^U, P_{\bar{g}_t \theta}^U)$$

Elevando al cuadrado:

$$\rho^2(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U) \leq 2 \rho^2(P_{\theta}^{\bar{g}_t \circ U}, P_{g_t^* \theta}^U) + 2 \rho^2(P_{g_t \theta}^U, P_{\bar{g}_t \theta}^U)$$

Estudiemos cada uno de los dos sumandos anteriores por separado. Por una parte y según (7.2), $\forall \gamma < \gamma_S$:

$$\lim_{t \rightarrow e} \frac{\rho^2(P_{g_t^* \theta}^U, P_{\theta}^{\bar{g}_t \circ U})}{\rho^2(g_t^* \theta, \theta)^\gamma} = 0$$

Respecto al otro término, observemos que $P_{g_t^* \theta}^U$ viene definida según:

$$P_{g_t^* \theta}^U(B) = P_{\theta}(U \circ g_t \in B) = P_{g_t^* \theta}(U^{-1}(B)), \quad B \in \mathcal{B}(\Theta)$$

y $P_{\bar{g}_t \theta}^U$ a través de:

$$P_{\bar{g}_t \theta}^U(B) = P_{\bar{g}_t \theta}(U \in B) = P_{\bar{g}_t \theta}(U^{-1}(B)), \quad B \in \mathcal{B}(\Theta)$$

Al variar B sobre todos los borelianos de Θ y ser U medible, podemos afirmar que $\rho(P_{\theta}^{U \circ g_t}, P_{\bar{g}_t \theta}^U) = \rho^2(g_t^* \theta, \bar{g}_t \theta)$. Pero conocemos que la familia \mathcal{P}_{Θ} es invariante de orden α_I con lo que $\forall \alpha < \alpha_I$:

$$\lim_{t \rightarrow e} \frac{\rho^2(P_{g_t^* \theta}^U, P_{\bar{g}_t \theta}^U)}{\rho^2(g_t^* \theta, \theta)^\alpha} = \lim_{t \rightarrow e} \frac{\rho^2(g_t^* \theta, \bar{g}_t \theta)}{\rho^2(g_t^* \theta, \theta)^\alpha} = 0$$

Por lo tanto, pese a que no podemos afirmar que la discrepancia entre $P_{\theta}^{\bar{v}_t \circ U}$ y $P_{\bar{v}_t \theta}^U$ sea 0, tiende a cero con una rapidez que podemos controlar respecto a $\rho^2(g_t^* \theta, \theta)$, esto es tiende a 0 en orden mayor o igual que $\gamma_S \wedge \alpha_I$.

El ejemplo anterior sugiere que pese a que los estimadores equivariantes en orden no son estables, si que lo serían en orden. Pasamos a concretar esta idea en la siguiente definición.

Supongamos que la discrepancia sea de clase \mathcal{C}^∞ y sea $\theta \in \Theta$. La diferenciabilidad del campo X permite definir una familia de aplicaciones de clase también \mathcal{C}^∞ :

$$\begin{aligned} I_\theta &\longrightarrow [0, +\infty) \\ t &\longmapsto \delta(P_{\theta}^{\bar{v}_t \circ U}, P_{\bar{v}_t \theta}^U) \end{aligned}$$

Definición 7.3.2. Diremos que el estimador U es estable de orden $\gamma_E \geq 1$ si y sólo si:

$$\gamma_E = \sup\{\gamma \geq 0 \mid \lim_{t \rightarrow e} \frac{\delta(P_{\theta}^{\bar{v}_t \circ U}, P_{\bar{v}_t \theta}^U)}{\delta(g_t^* \theta, \theta)^\gamma} = 0, \quad \forall \theta \in \Theta\}$$

Ejemplo 7.3.3. Volviendo al ejemplo 7.3.2, observemos que $\forall \theta \in \Theta$ y $\forall \gamma < \gamma_S \wedge \alpha_I$:

$$\lim_{t \rightarrow e} \frac{\rho^2(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U)}{\rho^2(P_{g_t^* \theta}, P_{\theta}^{\gamma})} = 0$$

De esta manera, U es un estimador estable de orden, como mínimo, $\gamma_S \wedge \alpha_I$.

Ejemplo 7.3.4. Observemos que el estimador constante no va a ser estable, ni estable en orden, en el sentido de la definición 7.3.2. Con las notaciones habituales, sea:

$$\begin{aligned} U : \quad \Omega^k &\longrightarrow \Theta \\ (w_1, \dots, w_k) &\longmapsto a \end{aligned}$$

donde $a \in \Theta$. Estamos interesados en comparar $P_{\theta}^{\bar{v}_t \circ U}$ con $P_{\bar{v}_t \theta}^U$, $\forall \theta \in \Theta$, y $t \in T$. La distancia de Hellinger al cuadrado hace posible analizar la discrepancia entre estas dos medidas de forma sencilla.

Observemos que $P_{\theta}^{\bar{v}_t \circ U} = \delta_{\bar{v}_t a}$, donde ahora δ denota la *delta de Kronecker*, y $P_{\bar{v}_t \theta}^U = \delta_a$. De esta manera, la distancia de Hellinger al cuadrado entre $P_{\theta}^{\bar{v}_t \circ U} = \delta_{\bar{v}_t a}$ y $P_{\bar{v}_t \theta}^U = \delta_a$ es máxima e igual a 2 $\forall \theta \in \Theta$, $t \in T$. Esta condición implica que U no es un estimador estable, ni estable en orden, tal y como queríamos verificar.

Analicemos con algo más de detalle la condición de estabilidad de un estimador U . Observemos que, fijado $\theta \in \Theta$, el criterio de estabilidad de un estimador U se basa en la comparación de dos medidas de probabilidad, $P_{\theta}^{\bar{v}_t \circ U}$ y $P_{\bar{v}_t \theta}^U$, en $(\Theta, \mathcal{B}(\Theta))$.

Supongamos que para los t considerados ambas distribuciones de probabilidad y P_{θ}^U sean absolutamente continuas respecto al volumen Riemanniano, que denotaremos v . Sean $h(\cdot; \theta, t)$ la densidad de $P_{\theta}^{\bar{v}_t \circ U}$, $g(\cdot; \theta, t)$ la densidad de $P_{\bar{v}_t \theta}^U$ y $f(\cdot; \theta)$ la densidad de P_{θ}^U todas ellas respecto a v . Observemos que $h(\gamma; \theta, e) = g(\gamma; \theta, e) = f(\gamma; \theta)$, $\forall \gamma \in \Theta$ y que:

$$h(\cdot; \theta, t), \quad g(\cdot; \theta, t), \quad \text{y} \quad f(\cdot; \theta) \in L^2(\Theta, \mathcal{B}(\Theta), v)$$

Identificaremos $P_{\theta}^{\bar{v}_t \circ U}$ y $P_{\bar{v}_t \theta}^U$ en $(\Theta, \mathcal{B}(\Theta))$ con dos veces las raíces cuadradas positivas de sus derivadas de Radon–Nikodym respecto a v .

En el entorno en el que ahora nos encontramos parece razonable tomar como Δ una subvariedad del espacio de Hilbert $L^2(\Theta, \mathcal{B}(\Theta), v)$, a poder ser de dimensión finita (por simplicidad, pero no es estrictamente necesario), y como discrepancia la distancia de Hellinger al cuadrado, d_H^2 . Según su definición:

$$d_H^2(P_{\theta}^{\overline{v}_t \circ U}, P_{\overline{v}_t \theta}^U) = \int_{\Theta} \left(\sqrt{h(\gamma; \theta, t)} - \sqrt{g(\gamma; \theta, t)} \right)^2 v(d\gamma)$$

Recordemos que en el lema 6.2.2 hemos demostrado que siempre que la curva $t \mapsto g_t^* \theta$ esté parametrizada por la longitud de arco (condición que podemos suponer):

$$\lim_{t \rightarrow e} \frac{\rho^2(g_t^* \theta, \theta)}{(t - e)^2} = 1$$

donde $\theta \in \Theta$ es arbitrario pero fijo. Así, nuestro objetivo según la definición 7.3.2, se ve reducido a analizar las derivadas en el origen de $d_H^2(P_{\theta}^{\overline{v}_t \circ U}, P_{\overline{v}_t \theta}^U)$. Necesitamos realizar un paso previo que se encuentra detallado en el siguiente resultado.

Proposición 7.3.1. *Con las notaciones de la sección, supongamos que valor e que hemos asociado a la identidad en nuestra familia de transformaciones es el 0.*

Entonces, las expansiones de Taylor de $t \mapsto h(\cdot; \theta, t)$ y de $t \mapsto g(\cdot; \theta, t)$ de orden 1 en el origen son:

$$g(\gamma; \theta, t) = f(\gamma; \theta) + \langle \text{grad}_2 f(\gamma; \overline{v}_t(\theta))|_{t=0}, V(\theta) \rangle_{\theta} \cdot t + O(t^2)$$

$$\begin{aligned} h(\gamma; \theta, t) = f(\gamma; \theta) - \left(\langle \text{grad}_1 f((\overline{v}_t)^{-1}(\gamma); \theta)|_{t=0}, V(\gamma) \rangle_{\gamma} + \right. \\ \left. + f(\gamma; \theta) \text{div}(V(\gamma)) \right) \cdot t + O(t^2) \end{aligned}$$

donde $V(\vartheta) = \frac{d}{dt} \overline{v}_t(\vartheta)|_{t=0}$, $\forall \vartheta \in \Theta$.

Demostración. Estudiemos, en primer lugar $h(\cdot; \theta, t)$. Observemos que $\overline{v}_t \circ T$ es la imagen de una ley con densidad. De esta manera podemos afirmar que, salvo en un conjunto de volumen Riemanniano 0:

$$h(\gamma; \theta, t) = f((\overline{v}_t)^{-1}(\gamma); \theta) \cdot |\det((\overline{v}_t)^{-1})'(\gamma)|, \quad \gamma \in \Theta \quad (7.3)$$

Por otra parte, y desarrollando en serie de Taylor respecto a t alrededor de $t = 0$, se verifica que:

$$\begin{aligned} h(\gamma; \theta, t) &= h(\gamma; \theta, 0) + \frac{d}{dt} h(\gamma; \theta, t)|_{t=0} \cdot t + O(t^2) = \\ &= f(\gamma; \theta) + \frac{d}{dt} h(\gamma; \theta, t)|_{t=0} \cdot t + O(t^2) \end{aligned}$$

Fijado $\gamma \in \Theta$, necesitamos conocer $\frac{d}{dt}h(\gamma; \theta, t)|_{t=0}$. Usando (7.3) y según la regla de la cadena, podemos escribir:

$$\begin{aligned} \frac{d}{dt}h(\gamma; \theta, t) &= \langle \text{grad}_1 f((\bar{v}_t)^{-1}(\gamma); \theta), \frac{d}{dt}(\bar{v}_t)^{-1}(\gamma) \rangle_{(\bar{v}_t)^{-1}(\gamma)} \cdot |\det((\bar{v}_t)^{-1})'(\gamma)| + \\ &\quad + f((\bar{v}_t)^{-1}(\gamma); \theta) \cdot \left(|\det((\bar{v}_t)^{-1})'(\gamma)| \right)' \end{aligned} \quad (7.4)$$

Estudiemos $\left(|\det((\bar{v}_t)^{-1})'(\gamma)| \right)'$. Nuevamente, desarrollando en serie de Taylor respecto a t alrededor de 0 obtenemos que:

$$((\bar{v}_t)^{-1})'(\gamma) = (v_0^{-1})'(\gamma) + \frac{d}{dt}((\bar{v}_t)^{-1})'(\gamma)|_{t=0} \cdot t + O(t^2)$$

Si ahora evaluamos el determinante de la matriz anterior obtenemos:

$$\begin{aligned} \det((\bar{v}_t)^{-1})'(\gamma) &= \prod_{\alpha=1}^n \left(1 + \frac{d}{dt} \frac{\partial}{\partial \gamma^\alpha} (\bar{v}_t)^{-1(\alpha)}(\gamma)|_{t=0} \cdot t + O(t^2) \right) + O(t^2) = \\ &= 1 + \sum_{\alpha=1}^n \left(\frac{d}{dt} \frac{\partial}{\partial \gamma^\alpha} (\bar{v}_t)^{-1(\alpha)}(\gamma)|_{t=0} \right) \cdot t + O(t^2) = \\ &= 1 + \text{tr} \left(\frac{d}{dt} ((\bar{v}_t)^{-1})'(\gamma)|_{t=0} \right) \cdot t + O(t^2) = \\ &= 1 + \text{div} \left(\frac{d}{dt} (\bar{v}_t)^{-1}(\gamma)|_{t=0} \right) \cdot t + O(t^2) \end{aligned}$$

donde $\bar{v}_t^{-1(\alpha)}$, $\alpha = 1, \dots, n$ son las funciones componentes de la aplicación \bar{v}_t^{-1} , $\text{tr}(\frac{d}{dt}((\bar{v}_t)^{-1})'(\gamma)|_{t=0})$ indica la traza de la matriz:

$$\frac{d}{dt}((\bar{v}_t)^{-1})'(\gamma)|_{t=0}$$

y $\text{div}(\frac{d}{dt}(\bar{v}_t)^{-1}(\gamma)|_{t=0})$ es la divergencia del campo vectorial $\frac{d}{dt}(\bar{v}_t)^{-1}(\gamma)|_{t=0}$. Sea V el campo vectorial en Θ definido según:

$$\Theta \ni \vartheta \longmapsto \frac{d}{dt} \bar{v}_t(\vartheta)|_{t=0}$$

Recuperando la expresión (7.4) y evaluando en $t = 0$ podemos escribir:

$$\frac{d}{dt}h(\gamma; \theta, t)|_{t=0} = \langle \text{grad}_1 f((\bar{v}_t)^{-1}(\gamma); \theta)|_{t=0}, -V(\gamma) \rangle_\gamma + f(\gamma; \theta) \text{div}(-V(\gamma))$$

con lo que:

$$\begin{aligned}
h(\gamma; \theta, t) = f(\gamma; \theta) - \left(\langle \text{grad}_1 f((\bar{v}_t)^{-1}(\gamma); \theta) \rangle_{|t=0}, V(\gamma) \rangle_\gamma + \right. \\
\left. + f(\gamma; \theta) \text{div}(V(\gamma)) \right) \cdot t + O(t^2)
\end{aligned} \tag{7.5}$$

Estudiemos ahora la función g . $\forall \gamma \in \Theta$, excepto quizás en un conjunto de volumen Riemanniano 0, se verifica:

$$g(\gamma; \theta, t) = f(\gamma; \bar{v}_t(\theta))$$

De esta manera, fijado γ y desarrollando nuevamente en serie de Taylor alrededor de $t = 0$:

$$g(\gamma; \theta, t) = f(\gamma; \theta) + \left(\frac{d}{dt} f(\gamma; \bar{v}_t(\theta)) \right)_{|t=0} \cdot t + O(t^2) \tag{7.6}$$

Necesitamos conocer $\frac{d}{dt} f(\gamma; \bar{v}_t(\theta))_{|t=0}$. De igual manera a como hicimos anteriormente, la Regla de la Cadena informa que:

$$\frac{d}{dt} f(\gamma; \bar{v}_t(\theta)) = \langle \text{grad}_2 f(\gamma; \bar{v}_t(\theta)), \frac{d}{dt} \bar{v}_t(\theta) \rangle_{\bar{v}_t(\theta)}$$

Evalutando en $t = 0$:

$$\frac{d}{dt} f(\gamma; \bar{v}_t(\theta))_{|t=0} = \langle \text{grad}_2 f(\gamma; \bar{v}_t(\theta))_{|t=0}, \frac{d}{dt} \bar{v}_t(\theta)_{|t=0} \rangle_\theta = \langle \text{grad}_2 f(\gamma; \theta), V(\theta) \rangle_\theta$$

Por lo tanto, si recuperamos la expresión (7.6) obtenemos:

$$g(\gamma; \theta, t) = f(\gamma; \theta) + \langle \text{grad}_2 f(\gamma; \bar{v}_t(\theta))_{|t=0}, V(\theta) \rangle_\theta \cdot t + O(t^2) \tag{7.7}$$

con lo que la proposición queda demostrada. □

Las expresiones (7.7) y (7.5) proporcionan las expansiones de Taylor de orden 1 en el origen de las densidades de $P_\theta^{\bar{v}_t \circ T}$ y de $P_{\bar{v}_t \theta}^T$ respecto al volumen Riemanniano. Observemos que por técnicas parecidas, aunque más laboriosas, se podrían obtener las expansiones de orden superior.

Volvamos ahora al problema de la estabilidad de U . La definición 7.3.2 establece que, si U es estable de orden mayor o igual que 2, la derivada primera de la aplicación:

$$I_\theta \ni t \mapsto d_H^2(P_\theta^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U)$$

debe ser nula en el origen.

Supongamos que se dan las condiciones de regularidad necesarias para poder intercambiar la esperanza y $\frac{d}{dt}$. Entonces:

$$\frac{d}{dt} d_H^2(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U)_{|t=0} = - \int_{\Theta} \frac{d}{dt} h(\gamma; \theta, t)_{|t=0} + \frac{d}{dt} g(\gamma; \theta, t)_{|t=0} v(d\gamma)$$

De esta manera, y según (7.5) y (7.7) podemos afirmar que:

$$\begin{aligned} 0 = \frac{d}{dt} d_H^2(P_{\theta}^{\bar{g}_t \circ U}, P_{\bar{g}_t \theta}^U)_{|t=0} &= \int_{\Theta} \left(\langle \text{grad}_1 f((\bar{v}_t)^{-1}(\gamma); \theta)_{|t=0}, V(\gamma) \rangle_{\gamma} + \right. \\ &\quad \left. + f(\gamma; \theta) \text{div}(V(\gamma)) \right) - \langle \text{grad}_2 f(\gamma; \bar{v}_t(\theta))_{|t=0}, V(\theta) \rangle_{\theta} v(d\gamma) \end{aligned}$$

Una posibilidad para que la expresión anterior se anule es que el integrando sea idénticamente 0. Sin embargo, en general la densidad respecto al volumen Riemanniano de la ley de un estimador estable de orden mayor o igual que 2 debe verificar:

$$\begin{aligned} \int_{\Theta} \langle \text{grad}_1 f((\bar{v}_t)^{-1}(\gamma); \theta)_{|t=0}, V(\gamma) \rangle_{\gamma} + f(\gamma; \theta) \text{div}(V(\gamma)) - \\ - \langle \text{grad}_2 f(\gamma; \bar{v}_t(\theta))_{|t=0}, V(\theta) \rangle_{\theta} v(d\gamma) = 0 \end{aligned} \quad (7.8)$$

Conocido el campo V , esta condición restringe la ley del estimador T . La solución, f , de esta ecuación es la densidad de P_{θ}^T respecto al volumen Riemanniano. El problema es que en general no conocemos el campo V sino que suponemos que existe una familia de transformaciones regulares (las transformaciones \bar{v}_t son localmente 1-paramétricas y localmente invertibles) de las que desconocemos su origen suponemos que son. Si quisiéramos que esta ecuación se verificara para cualquier campo V difícilmente encontraríamos la solución. Es en este punto donde toman importancia los campos de Killing. Observemos que si X es un campo de este tipo, la aplicación:

$$I_{\theta} \ni t \mapsto v_t(\theta) \in \Delta$$

es una isometría local. Además, es un resultado conocido en el ámbito de la Geometría Diferencial que el conjunto de campos de Killing de una variedad Riemanniana forma un álgebra de Lie de dimensión menor o igual que $\frac{(m+1)m}{2}$, donde m es la dimensión de la variedad. Una demostración detallada al respecto puede consultarse en la página 165 de Petersen [33].

De esta manera, imponiendo que la ecuación (7.8) se verifique para cada uno de los elementos de la base de este álgebra de Lie, dispondríamos de, a lo

sumo, $\frac{(m+1)m}{2}$ ecuaciones en derivadas parciales cuya solución $f(\cdot; \theta)$ sería la de aquellos estimadores estables de orden mayor o igual que 2.

Sin embargo, el campo:

$$\Theta \ni \vartheta \longmapsto V(\vartheta) = \frac{d}{dt} \overline{v}_t(\vartheta)|_{t=0}$$

en general no es de Killing. Vamos a ver a continuación la relación entre el campo anterior y:

$$\Theta \ni \vartheta \longmapsto \frac{d}{dt} v_t(\vartheta)|_{t=0}$$

que sí podemos suponer de Killing.

Antes de continuar debemos comentar que otras clases razonables de campos a considerar. Podríamos tomar campos contruidos como la suma de uno de Killing en Θ y una contaminación o perturbación en Δ . También podría ser interesante analizar transportes paralelos de vectores constantes a lo largo de una geodésica de Θ . Esperamos, en futuras investigaciones, clarificar el papel que desempeñan este tipo de campos en la estimación estable.

Proposición 7.3.2. *Sea Δ una variedad Riemanniana m -dimensional y Θ una subvariedad de Δ de dimensión n , $m \geq n + 1$. Sean δ la discrepancia escogida en Δ y ρ la distancia Riemanniana en Δ .*

Entonces, $\forall \vartheta \in \Theta$:

$$\frac{d}{dt} v_t(\vartheta)|_{t=0}$$

es la proyección, según la distancia definida a través de la métrica Riemanniana, de:

$$\frac{d}{dt} \overline{v}_t(\vartheta)|_{t=0}$$

en el espacio $T_{\vartheta}(\Theta)$.

Demostración. En primer lugar, recordemos que para la familia de discrepancias δ que consideramos se verifica (con la notación de Einstein o de índices repetidos):

$$\delta(\vartheta, \vartheta + \Delta \vartheta) = c \cdot g_{\alpha, \beta}(\vartheta) \Delta \vartheta_{\alpha} \Delta \vartheta_{\beta} + O(\Delta \vartheta^3), \quad \vartheta \in \Delta$$

siendo $(g_{\alpha, \beta}(\vartheta))_{\alpha, \beta=1, \dots, m}$ la matriz de la métrica en el punto ϑ y c una constante. Observemos que la distancia de Hellinger al cuadrado verifica la condición anterior.

Sea ahora $\theta \in \Theta \subset \Delta$. Consideremos dos sistemas de coordenadas locales en θ dados por:

$$\begin{aligned} x: V \subset \Delta &\longrightarrow \mathbb{R}^m \\ p &\longmapsto (x^1(p), \dots, x^m(p)) \\ \\ u: U \subset \Delta &\longrightarrow \mathbb{R}^n \\ \vartheta &\longmapsto (u^1(\vartheta), \dots, u^n(\vartheta)) \end{aligned}$$

Sea $\vartheta \in U$ y denotemos:

$$\begin{aligned} u^1(\vartheta) &= u^1, \dots, u^n(\vartheta) = u^n \\ u^1(\theta) &= u_0^1, \dots, u^n(\theta) = u_0^n \end{aligned}$$

Sea $\underline{x}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ definida según $\underline{x} = x \circ u^{-1}$. Observemos que un punto $\vartheta \in U$ puede escribirse en coordenadas como:

$$(x \circ u^{-1})(u^1, \dots, u^n) = \underline{x}(u^1, \dots, u^n) = (x^1(u^1, \dots, u^n), \dots, x^m(u^1, \dots, u^n))$$

Recordemos que estamos considerando una perturbación de $\theta \in \Theta$ a través de:

$$\begin{aligned} I_\theta &\longrightarrow \Delta \\ t &\longmapsto v_t(\theta) \end{aligned}$$

Si escribimos los puntos de la curva anterior en coordenadas:

$$\begin{aligned} I_\theta &\rightarrow \Delta \rightarrow \mathbb{R}^m \\ t &\mapsto v_t(\theta) \mapsto x(v_t(\theta)) \end{aligned}$$

Como $\theta \in \Theta$ es fijo, escribamos $x(v_t(\theta)) = x(t)$ y observemos que $x(0) = x(\theta) = \underline{x}(u_0)$.

Fijado $t \in I_\theta$, estamos interesados en un elemento de $\mathcal{P}_\Theta = \{P_\vartheta \mid \vartheta \in \Theta\}$, que denotaremos $P_{\theta'}$ para el que:

$$\delta(v_t(\theta), P_{\theta'}) = \min_{\vartheta \in \Theta} \delta(v_t(\theta), P_\vartheta)$$

Como t está próximo a 0, podemos identificar $P_{\theta'}$ y P_ϑ con sus etiquetas pues recordemos que el modelo es localmente identificable. De esta manera, escribiendo los puntos de Θ y de Δ en los sistemas de coordenadas introducidos anteriormente, $u(\theta')$ debe obtenerse como solución del sistema:

$$\frac{\partial}{\partial u^\alpha} \delta(x(t), \underline{x}(u)) = 0, \quad \alpha = 1, \dots, n$$

Sin embargo, las ecuaciones anteriores vistas como función de $t \in I_\theta$ son una constante, por lo que también debe verificarse:

$$\frac{d}{dt} \frac{\partial}{\partial u^\alpha} \delta(x(t), \underline{x}(u)) = 0, \quad \alpha = 1, \dots, n \quad (7.9)$$

Explicitemos este sistema. Recordemos que δ verifica:

$$\delta(x(t), \underline{x}(u)) = c \cdot g_{ij}(x(t)) (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) + O(\Delta(t)^3)$$

donde, con abuso de notación, $\Delta(t)^3$ indica términos que contienen $(\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t))$. Sin pérdida de generalidad vamos a suponer que $c = 1$. Con esta información, las ecuaciones de (7.9) se escriben:

$$\begin{aligned} 0 &= \frac{d}{dt} \frac{\partial}{\partial u^\alpha} \delta(x(t), \underline{x}(u)) = \frac{\partial}{\partial u^\alpha} \left(g_{ij}(x(t)) (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) \right) + \\ &\quad + \frac{d}{dt} \frac{\partial}{\partial u^\alpha} \left(O(\Delta(t)^3) \right) \end{aligned}$$

donde $\alpha \in \{1, \dots, n\}$. Estudiemos por separado cada uno de los dos sumandos anteriores.

$$\begin{aligned} &\frac{\partial}{\partial u^\alpha} \left(g_{ij}(x(t)) (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) \right) = \\ &= g_{ij}(x(t)) \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) (\underline{x}^j(u) - x^j(t)) + g_{ij}(x(t)) (\underline{x}^i(u) - x^i(t)) \frac{\partial}{\partial u^\alpha} \underline{x}^j(u) = \\ &= 2 g_{ij}(x(t)) \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) (\underline{x}^j(u) - x^j(t)) \end{aligned}$$

pues $g_{ij}(x(t)) = g_{ji}(x(t))$, $\forall i, j \in \{1, \dots, m\}$. Si ahora derivamos respecto a t :

$$\begin{aligned} &\frac{d}{dt} \frac{\partial}{\partial u^\alpha} \left(g_{ij}(x(t)) (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) \right) = \\ &= \frac{d}{dt} \left(2 g_{ij}(x(t)) \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) (\underline{x}^j(u) - x^j(t)) \right) \\ &= 2 \left(\left(\frac{d}{dt} g_{ij}(x(t)) \right) \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) (\underline{x}^j(u) - x^j(t)) + \right. \end{aligned}$$

$$\begin{aligned}
& +g_{ij}(x(t)) \left(\frac{d}{dt} \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) \right) (\underline{x}^j(u) - x^j(t)) + \\
& +g_{ij}(x(t)) \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) \left(\frac{\partial}{\partial u^\beta} \underline{x}^j(u) \frac{d}{dt} u^\beta - \frac{d}{dt} x^j(t) \right)
\end{aligned}$$

Estudiemos ahora la parte correspondiente a $\frac{d}{dt} \frac{\partial}{\partial u^\alpha} (O(\Delta(t)^3))$.

Para nuestros propósitos, bastará que analicemos $\frac{\partial}{\partial u^\alpha} (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t))$. Se verifica que:

$$\begin{aligned}
& \frac{\partial}{\partial u^\alpha} \left((\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t)) \right) = \\
& = 3 \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t))
\end{aligned}$$

Derivando la expresión anterior respecto a t :

$$\begin{aligned}
& \frac{\partial}{\partial u^\alpha} \left(\frac{d}{dt} (\underline{x}^i(u) - x^i(t)) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t)) \right) = \\
& = 3 \left(\frac{d}{dt} \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) \right) (\underline{x}^j(u) - x^j(t)) (\underline{x}^k(u) - x^k(t)) + \\
& + 6 \frac{\partial}{\partial u^\alpha} \underline{x}^i(u) \left(\frac{d}{dt} (\underline{x}^j(u) - x^j(t)) \right) (\underline{x}^k(u) - x^k(t))
\end{aligned}$$

Nosotros estamos interesados en relacionar los vectores tangentes a las curvas:

$$\begin{aligned}
I_\theta & \longrightarrow \Delta \\
t & \longmapsto v_t(\theta)
\end{aligned}$$

y:

$$\begin{aligned}
I_\theta & \longrightarrow \Delta \\
t & \longmapsto \bar{v}_t(\theta)
\end{aligned}$$

en el origen. Por lo tanto, evaluemos $\frac{d}{dt} \frac{\partial}{\partial u^\alpha} \rho^2(x(t), \underline{x}(u))$ en $t = 0$ y en $u = u_0$. Se tiene que:

$$\begin{aligned}
0 & = \frac{d}{dt} \frac{\partial}{\partial u^\alpha} \rho^2(x(t), \underline{x}(u))|_{t=0, u=u_0} = \\
& = \frac{\partial}{\partial u^\alpha} \underline{x}^i(u)|_{u=u_0} g_{ij}(u_0) \frac{\partial}{\partial u^\beta} \underline{x}^j(u)|_{u=u_0} \frac{du^\beta}{dt} \Big|_{t=0} -
\end{aligned}$$

$$-\frac{\partial}{\partial u^\alpha} \underline{x}^i(u)|_{u=u_0} g_{ij}(u_0) \frac{d}{dt} x^j(t)|_{t=0}, \quad \forall \alpha = 1, \dots, n \quad (7.10)$$

pues recordemos que $x(0) = x(\theta) = \underline{x}(u_0)$ y por lo tanto los términos $O(\Delta(t)^3)$ no contribuyen.

Podemos escribir las ecuaciones anteriores en forma matricial. Sean:

$$\mathbf{X} = \begin{pmatrix} \frac{\partial}{\partial u^1} \underline{x}^1(u)|_{u=u_0} & \dots & \frac{\partial}{\partial u^n} \underline{x}^1(u)|_{u=u_0} \\ \vdots & & \vdots \\ \frac{\partial}{\partial u^1} \underline{x}^m(u)|_{u=u_0} & \dots & \frac{\partial}{\partial u^n} \underline{x}^m(u)|_{u=u_0} \end{pmatrix}$$

$\mathbf{G}(u_0) = (g_{ij}(u_0))_{i,j=1,\dots,m}$ y:

$$\frac{du}{dt}|_{t=0} = \begin{pmatrix} \frac{du^1}{dt}|_{t=0} \\ \vdots \\ \frac{du^n}{dt}|_{t=0} \end{pmatrix} \quad \frac{dx}{dt}|_{t=0} = \begin{pmatrix} \frac{dx^1}{dt}|_{t=0} \\ \vdots \\ \frac{dx^m}{dt}|_{t=0} \end{pmatrix}$$

Entonces, las ecuaciones de (7.10) se escriben en forma matricial:

$$\mathbf{X}' \mathbf{G}(u_0) \mathbf{X} \frac{du}{dt}|_{t=0} = \mathbf{X}' \mathbf{G}(u_0) \frac{dx}{dt}|_{t=0}$$

Estas ecuaciones informan que el vector $\frac{du}{dt}|_{t=0}$ es la única proyección ortogonal de $\frac{dx}{dt}|_{t=0}$ en el subespacio $\{\mathbf{X} x \mid x \in \mathbb{R}^n\}$ de \mathbb{R}^m según la distancia definida en \mathbb{R}^m a través del producto escalar:

$$\langle u, v \rangle = u' \mathbf{G}(u_0) v, \quad u, v \in \mathbb{R}^m$$

Observemos que las columnas de \mathbf{X} son los elementos de la base, en coordenadas, del espacio $T_\theta(\Theta)$. Como todo este razonamiento es independiente de los sistemas de coordenadas escogidos, podemos afirmar que $\frac{d}{dt} \overline{v}_t(\theta)|_{t=0}$ es la proyección ortogonal de $\frac{d}{dt} v_t(\theta)|_{t=0}$ en $T_\theta(\Theta)$ según la distancia definida en $T_\theta(\Delta)$ a través del producto escalar:

$$\langle u, v \rangle = u' \mathbf{G}(\theta) v, \quad u, v \in T_\theta(\Delta) \quad (7.11)$$

tal y como queríamos demostrar. □

Una vez relacionados los vectores tangentes $\frac{d}{dt} \overline{v}_t(\theta)|_{t=0}$ y $\frac{d}{dt} v_t(\theta)|_{t=0}$ recuperemos la ecuación (7.8):

$$\begin{aligned}
\int_{\Theta} \langle \text{grad}_1 f((\overline{v}_t)^{-1}(\gamma); \theta), V(\gamma) \rangle_{\gamma} + f(\gamma; \theta) \text{div}(V(\gamma)) v(d\gamma) = \\
= \int_{\Theta} \langle \text{grad}_2 f(\gamma; \theta), V(\theta) \rangle_{\theta} v(d\gamma)
\end{aligned}$$

en la que conocido $V(\theta) = \frac{d}{dt} \overline{v}_t(\theta)|_{t=0}$, la solución f de esta ecuación es la densidad de P_{θ}^T , la ley de un estimador T estable de orden mayor o igual que 2, respecto al volumen Riemanniano.

Observemos que cualquier estimador que sea estable lo será de orden como mínimo 2, gracias a la suavidad de las perturbaciones.

Consideremos ahora un campo de Killing X . Conocemos que el conjunto de los campos de Killing en una variedad tiene dimensión menor o igual que $\frac{m(m+1)}{2}$, donde m es la dimensión de la variedad. Sea $\theta \in \Theta$ y proyectemos ortogonalmente $X(\theta)$ sobre $T_{\theta}(\Theta)$, según la distancia definida a través del producto escalar de (7.11). La proyección resultante es el vector tangente $V(\theta)$ que requeríamos anteriormente y la solución de las, a lo sumo, $\frac{m(m+1)}{2}$ e.d.p.'s que obtendríamos de (7.8), la densidad respecto al volumen Riemanniano que buscamos. El siguiente ejemplo ilustra este procedimiento.

Ejemplo 7.3.5. Sea X un campo de Killing en una variedad Riemanniana Δ . Sea g la métrica en Δ . Por definición de campo de Killing, X debe verificar:

$$(L_X g)(Y, Z) = 0, \quad \forall Y, Z \text{ campos } \mathcal{C}^{\infty} \text{ en } \Delta$$

donde L_X es la derivada de Lie vía X . Detallemos un poco más la condición anterior:

$$\begin{aligned}
0 &= (L_X g)(Y, Z) = X g(Y, Z) - g([X, Y], Z) - g(Y, [X, Z]) = \\
&= g(D_X Y, Z) + g(Y, D_X Z) - g([X, Y], Z) - g(Y, [X, Z]) = \\
&= g(D_X Y - [X, Y], Z) + g(Y, D_X Z - [X, Z]) = \\
&= \langle D_X Y - [X, Y], Z \rangle + \langle Y, D_X Z - [X, Z] \rangle
\end{aligned}$$

siendo $[X, Y]$ el paréntesis de Lie entre X e Y , D la conexión Riemanniana o de Levi-Civita en Δ y $D_X Y$ la derivada covariante de Y en la dirección de X .

Expresemos este último término en coordenadas. Sea $(\theta^1, \dots, \theta^m)$ un sistema de coordenadas en Δ . Nuevamente con la notación de índices repetidos:

$$D_X Y = \left(\frac{\partial Y^i}{\partial \theta^\alpha} X^\alpha + \Gamma_{\alpha\beta}^i Y^\alpha X^\beta \right) \frac{\partial}{\partial \theta^i}$$

donde $(\Gamma_{\alpha\beta}^i)_{i,\alpha,\beta=1,\dots,m}$ son los símbolos de Christoffel de segunda especie correspondientes a D y:

$$[X, Y] = \left(X^j \frac{\partial Y^i}{\partial \theta^j} - Y^j \frac{\partial X^i}{\partial \theta^j} \right) \frac{\partial}{\partial \theta^i}$$

De esta manera:

$$D_X Y - [X, Y] = \left(\frac{\partial X^i}{\partial \theta^\alpha} Y^\alpha + \Gamma_{\alpha\beta}^i Y^\alpha X^\beta \right) \frac{\partial}{\partial \theta^i}$$

$$D_X Z - [X, Z] = \left(\frac{\partial X^i}{\partial \theta^\alpha} Z^\alpha + \Gamma_{\alpha\beta}^i Z^\alpha X^\beta \right) \frac{\partial}{\partial \theta^i}$$

Realicemos ahora los productos escalares. Por una parte:

$$\langle D_X Y - [X, Y], Z \rangle = g_{ik} \left(\frac{\partial X^i}{\partial \theta^\alpha} Y^\alpha + \Gamma_{\alpha\beta}^i Y^\alpha X^\beta \right) Z^k$$

mientras que:

$$\langle Y, D_X Z - [X, Z] \rangle = g_{ik} \left(\frac{\partial X^i}{\partial \theta^\alpha} Z^\alpha + \Gamma_{\alpha\beta}^i Z^\alpha X^\beta \right) Y^k$$

Sumando ambas expresiones:

$$\begin{aligned} & \langle D_X Y - [X, Y], Z \rangle + \langle Y, D_X Z - [X, Z] \rangle = \\ & = g_{ik} \frac{\partial X^i}{\partial \theta^\alpha} (Y^\alpha Z^k + Y^k Z^\alpha) + [\alpha\beta, k] Y^\alpha X^\beta Z^k + [\alpha\beta, k] Z^\alpha X^\beta Y^k \end{aligned}$$

donde $([\alpha\beta, k])_{\alpha,\beta,k=1,\dots,m}$ son los símbolos de Christoffel de primera especie. Observemos que:

$$\begin{aligned} & [\alpha\beta, k] Y^\alpha X^\beta Z^k + [\alpha\beta, k] Z^\alpha X^\beta Y^k = \\ & = \frac{1}{2} \left(\frac{\partial g_{\alpha k}}{\partial \theta^\beta} + \frac{\partial g_{\beta k}}{\partial \theta^\alpha} + \frac{\partial g_{\alpha\beta}}{\partial \theta^k} \right) Y^\alpha X^\beta Z^k + \\ & + \frac{1}{2} \left(\frac{\partial g_{\alpha k}}{\partial \theta^\beta} + \frac{\partial g_{\beta k}}{\partial \theta^\alpha} + \frac{\partial g_{\alpha\beta}}{\partial \theta^k} \right) Z^\alpha X^\beta Y^k = \frac{\partial g_{\alpha k}}{\partial \theta^\beta} Y^\alpha X^\beta Z^k \end{aligned}$$

De esta manera, $\forall Y, Z$ debe verificarse:

$$(L_X g)(Y, Z) = Y^\alpha \left(g_{ik} \frac{\partial X^i}{\partial \theta^\alpha} + g_{i\alpha} \frac{\partial X^i}{\partial \theta^k} + \frac{\partial g_{\alpha k}}{\partial \theta^\beta} X^\beta \right) Z^k = 0$$

esto es:

$$g_{ik} \frac{\partial X^i}{\partial \theta^\alpha} + g_{i\alpha} \frac{\partial X^i}{\partial \theta^k} + \frac{\partial g_{\alpha k}}{\partial \theta^\beta} X^\beta = 0, \quad \forall k, \alpha = 1, \dots, m \quad (7.12)$$

Veamos que aspecto toman estas ecuaciones para la familia normal univariante. Recordemos que, en este caso, la matriz de la métrica es:

$$\mathbf{G}(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

Con esta información, las ecuaciones (7.12) se escriben:

$$\begin{aligned} \frac{2}{\sigma^2} \frac{\partial X^1}{\partial \mu} - \frac{2}{\sigma^3} X^2 &= 0, & k = \alpha = 1 \\ \frac{1}{\sigma^2} \frac{\partial X^1}{\partial \sigma} + \frac{2}{\sigma^2} \frac{\partial X^2}{\partial \mu} &= 0, & k = 1, \alpha = 2 \\ \frac{2}{\sigma^2} \frac{\partial X^2}{\partial \mu} + \frac{1}{\sigma^2} \frac{\partial X^1}{\partial \sigma} &= 0, & k = 2, \alpha = 2 \\ \frac{4}{\sigma^2} \frac{\partial X^2}{\partial \sigma} - \frac{4}{\sigma^3} X^2 &= 0, & k = \alpha = 2 \end{aligned}$$

Observemos que las ecuaciones segunda y la tercera son la misma. De esta manera, disponemos de 3 ecuaciones diferentes a solucionar. Destaquemos que 3 coincide con el valor anunciado de $\frac{m(m+1)}{2}$ pues en este caso, la dimensión de $\Delta = \mathbb{R} \times \mathbb{R}_+$ es 2.

Para la ecuación correspondiente a $k = \alpha = 2$:

$$\frac{4}{\sigma^2} \frac{\partial X^2}{\partial \sigma} - \frac{4}{\sigma^3} X^2 = 0 \iff \frac{1}{X^2} \frac{\partial X^2}{\partial \sigma} = \frac{1}{\sigma} \iff X^2(\mu, \sigma) = A(\mu) \sigma$$

donde $A(\mu)$ es una función diferenciable a determinar que depende de μ . Si ahora analizamos las ecuaciones correspondientes a $k = 1, \alpha = 2$ y a $k = \alpha = 1$, tenemos que:

$$2 \frac{\partial X^2}{\partial \mu} + \frac{\partial X^1}{\partial \sigma} = 0 \iff 2 A'(\mu) \sigma + \frac{\partial X^1}{\partial \sigma} = 0$$

y:

$$\frac{\partial X^1}{\partial \mu} - A(\mu) = 0 \iff X^1 = \int A(\mu) d\mu + B(\sigma)$$

donde $B(\sigma)$ es una función diferenciable que depende de σ a determinar. Recopilando la información toda la información, obtenemos que:

$$2 A'(\mu) \sigma + B'(\sigma) = 0, \quad \forall (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$$

Una posibilidad sería que $A'(\mu)$ y $B'(\sigma)$ fuesen cero. En tal caso, sean $a = A(\mu)$, $\forall \mu \in \mathbb{R}$ y $b = B(\sigma)$, $\forall \sigma \in \mathbb{R}_+$. Entonces:

$$X^1(\mu, \sigma) = a \mu + b$$

$$X^2(\mu, \sigma) = a \sigma$$

Por otra parte, podría ser que $A'(\mu)$ y $B'(\sigma)$ no se anularan. De esta manera, debería verificarse:

$$2 A'(\mu) = -\frac{B'(\sigma)}{\sigma} = \gamma, \quad \forall (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$$

donde γ es una constante. Por lo tanto, $A(\mu) = \frac{\gamma}{2}\mu$, $\mu \in \mathbb{R}$ y $B(\sigma) = -\frac{\gamma}{2}\sigma^2$, $\sigma \in \mathbb{R}_+$. Sea $c = \frac{\gamma}{2}$. Entonces:

$$X^1(\mu, \sigma) = c \left(\frac{\mu^2}{2} - \sigma^2 \right)$$

$$X^2(\mu, \sigma) = c \mu \sigma$$

De esta manera, los campos de Killing para la familia Normal univariante son:

$$\begin{pmatrix} X^1(\mu, \sigma) \\ X^2(\mu, \sigma) \end{pmatrix} = a \begin{pmatrix} \mu \\ \sigma \end{pmatrix} + b \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} \frac{\mu^2}{2} - \sigma^2 \\ \mu \sigma \end{pmatrix}, \quad a, b, c \in \mathbb{R}$$

Centrémonos ahora en la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$. Observemos que en este caso Θ es la variedad 1-dimensional $\mathbb{R} \times \{1\}$. Sea $\mu \in \mathbb{R}$. Según la expresión anterior, el conjunto de los campos de Killing en $T_{(\mu, 1)}(\Delta) = T_{(\mu, 1)}(\mathbb{R} \times \mathbb{R}_+)$ viene dado por:

$$\begin{pmatrix} X^1(\mu, \sigma) \\ X^2(\mu, \sigma) \end{pmatrix} = a \begin{pmatrix} \mu \\ 1 \end{pmatrix} + b \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad a, b \in \mathbb{R}$$

Observemos que:

$$\begin{pmatrix} \mu \\ 1 \end{pmatrix} \text{ y } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad a, b \in \mathbb{R}$$

son linealmente independientes para cualquier valor de μ . De esta manera, estamos interesados en encontrar la proyección de cada uno de ellos sobre $T_{(\mu,1)}(\Theta)$ según la distancia definida a través de:

$$\mathbf{G}(\mu, 1) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

esto es, busquemos α y $\beta \in \mathbb{R}$ de manera que:

$$(\mu - \alpha \quad 1) \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \mu - \alpha \\ 1 \end{pmatrix} = (\mu - \alpha)^2 + 2$$

y:

$$(1 - \beta \quad 0) \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 - \beta \\ 0 \end{pmatrix} = (1 - \beta)^2$$

sean mínimas. Es inmediato verificar que en el primer caso, la solución es $\mu = \alpha$ y en el segundo $\beta = 1$.

De esta manera, las proyecciones buscadas de $\begin{pmatrix} \mu - \alpha \\ 1 \end{pmatrix}$ y de $\begin{pmatrix} 1 - \beta \\ 0 \end{pmatrix}$ sobre $T_{(\mu,1)}(\Theta)$ son:

$$V_1(\mu, 1) = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \text{ y } V_2(\mu, 1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

respectivamente. Veamos que transformaciones definen V_1 y V_2 en Θ . Vamos a imponer que las curvas integrales de uno y otro campo verifiquen una

condición adicional: requeriremos de ellas que sean geodésicas, esto es que si T es el vector tangente a la curva, la derivada covariante $D_T T = 0$, donde D es la conexión Riemanniana.

Estudiemos primero las curvas integrales de V_1 . Fijado $\mu \in \mathbb{R}$, estamos interesados en una curva $t \mapsto (\overline{v_{1,t}}(\mu, 1), 1) \in \Theta$ de manera que:

$$\frac{d}{dt} \overline{v_{1,t}}(\mu, 1) = \mu$$

Así, $\overline{v_{1,t}}(\mu, 1) = t\mu + c$, donde c es una constante a determinar. La condición inicial informa que $\overline{v_{1,1}}(\mu, 1) = \mu$. De esta manera, $c = 0$ y la curva que obtenemos en Θ es:

$$\begin{aligned} \mathbb{R} &\longrightarrow \Theta = \{(\mu, 1) \mid \mu \in \mathbb{R}\} \\ t &\mapsto (t\mu, 1) \end{aligned}$$

Observemos que coincide con la que obteníamos en el ejemplo 3.2.1 al perturbar la familia $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ mediante del grupo multiplicativo de los reales positivos $\mathcal{G} = \{g_t \mid t > 0\}$ (donde $g_t x = tx$, $x \in \mathbb{R}$) y a continuación proyectar en $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ para encontrar el elemento de esta familia que estuviera a distancia cuadrática de Rao mínima de la distribución $N(t\mu, t)$. Respecto a V_2 , buscamos una curva $t \mapsto (\overline{v_{2,t}}(\mu, 1), 1) \in \Theta$ tal que:

$$\frac{d}{dt} \overline{v_{2,t}}(\mu, 1) = 1$$

Debe ser $\overline{v_{2,t}}(\mu, 1) = t + c$, donde nuevamente c es una constante a determinar. La condición inicial establece que $\overline{v_{2,0}}(\mu, 1) = \mu$ con lo que $c = \mu$ y la nueva curva en Θ es:

$$\begin{aligned} \mathbb{R} &\longrightarrow \Theta = \{(\mu, 1) \mid \mu \in \mathbb{R}\} \\ t &\mapsto (t + \mu, 1) \end{aligned}$$

Nuevamente obtenemos un resultado conocido. Recordemos que la familia de distribuciones de probabilidad $\{N(\mu, 1) \mid \mu \in \mathbb{R}\}$ es invariante (de orden infinito) bajo la acción del grupo aditivo real $\mathcal{G} = \{g_t \mid t \in \mathbb{R}\}$ (donde $g_t x = t + x$, $x \in \mathbb{R}$). Fijado un elemento $g_t \in \mathcal{G}$, esta invariancia en sentido clásico permite definir una familia de transformaciones de Θ en Θ dadas por:

$$\begin{aligned} \Theta &\longrightarrow \Theta \\ (\mu, 1) &\mapsto (t + \mu, 1) \end{aligned}$$

de donde, fijado $(\mu, 1) \in \Theta$ y variando $t \in \mathbb{R}$, se tiene la curva $t \mapsto (\overline{v_{2,t}}(\mu, 1), 1)$.

Veamos ahora el aspecto que toma la ecuación (7.8):

$$\int_{\Theta} \langle \text{grad}_1 f((\overline{v_t})^{-1}(\gamma); \theta)|_{t=0}, V(\gamma) \rangle_{\gamma} + f(\gamma; \theta) \text{div}(V(\gamma)) - \\ - \langle \text{grad}_2 f(\gamma; \overline{v_t}(\theta))|_{t=0}, V(\theta) \rangle_{\theta} v(d\gamma) = 0$$

en nuestro caso particular.

Para $V_1(\mu, 1) = \begin{pmatrix} \mu \\ 0 \end{pmatrix}$ la condición anterior se escribe:

$$\int_{\Theta} \frac{\partial}{\partial \gamma} f(\gamma; \mu) \gamma + f(\gamma; \mu) - \frac{\partial}{\partial \mu} f(\gamma; \mu) \mu v(d\gamma) = 0 \quad (7.13)$$

y para $V_2(\mu, 1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$:

$$\int_{\Theta} \frac{\partial}{\partial \gamma} f(\gamma; \mu) - \frac{\partial}{\partial \mu} f(\gamma; \mu) v(d\gamma) = 0 \quad (7.14)$$

Veamos que la ley del estimador \overline{X}_n , esto es de la media basada en n observaciones independientes, verifica estas dos ecuaciones. Conocemos que \overline{X}_n se distribuye según una ley $N(\mu, \frac{1}{\sqrt{n}})$. De esta manera, debemos probar que:

$$f(x; \mu) = \frac{1}{2\pi \frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi \frac{1}{\sqrt{n}}}(x-\mu)^2}, \quad x \in \mathbb{R}$$

verifica (7.13) y (7.14). Observemos que:

$$\frac{\partial}{\partial x} f(x; \mu) = \frac{1}{2\pi \frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi \frac{1}{\sqrt{n}}}(x-\mu)^2} (-n(x-\mu))$$

$$\frac{\partial}{\partial \mu} f(x; \mu) = \frac{1}{2\pi \frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi \frac{1}{\sqrt{n}}}(x-\mu)^2} n(x-\mu)$$

De esta manera, (7.13) se escribe:

$$\int_{\mathbb{R}} \frac{1}{2\pi \frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi \frac{1}{\sqrt{n}}}(x-\mu)^2} (-n(x-\mu)) + \frac{1}{2\pi \frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi \frac{1}{\sqrt{n}}}(x-\mu)^2} -$$

$$\begin{aligned}
& -\frac{1}{2\pi\frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi\frac{1}{\sqrt{n}}}(x-\mu)^2} n(x-\mu) dx = \\
& = 1 - n \int_{\mathbb{R}} \frac{1}{2\pi\frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi\frac{1}{\sqrt{n}}}(x-\mu)^2} (x^2 - \mu^2) dx = 1 - n \left(\frac{1}{n} + \mu^2 - \mu^2 \right) = 0
\end{aligned}$$

tal y como queríamos ver. Respecto a la expresión (7.13):

$$\begin{aligned}
& \int_{\mathbb{R}} \frac{1}{2\pi\frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi\frac{1}{\sqrt{n}}}(x-\mu)^2} (-n(x-\mu)) - \frac{1}{2\pi\frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi\frac{1}{\sqrt{n}}}(x-\mu)^2} n(x-\mu) dx = \\
& = -2n \int_{\mathbb{R}} \frac{1}{2\pi\frac{1}{\sqrt{n}}} \exp^{-\frac{1}{2\pi\frac{1}{\sqrt{n}}}(x-\mu)^2} (x-\mu) dx = 0
\end{aligned}$$

De esta manera, queda demostrado que $\overline{X_n}$ es un estimador estable de orden, por lo menos 2. Este resultado concuerda con las ideas detalladas en el ejemplo 7.3.3 pues aquí el orden de la familia α_I es 1, el del estimador γ_S es $+\infty$ y por lo tanto $2(\gamma_S \wedge \alpha_I) = 2$

Observemos que cualquier estimador equivariante bajo la acción del grupo afín debe verificar la condición (7.14) por su propia construcción. Podemos comprobarlo recordando que las leyes de estos estimadores deben ser de la forma $f(x; \mu) = F(x - \mu)$, $x \in \mathbb{R}$, donde F es una función conveniente. Así:

$$\frac{\partial}{\partial x} f(x; \mu) = F'(x - \mu), \quad \frac{\partial}{\partial \mu} f(x; \mu) = -F'(x - \mu)$$

con lo que:

$$\int_{\Theta} \frac{\partial}{\partial x} f(x; \mu) - \frac{\partial}{\partial \mu} f(x; \mu) dx = 2 \int_{\mathbb{R}} \frac{\partial}{\partial \mu} f(x; \mu) dx = 0$$

pues por definición de densidad $\int_{\mathbb{R}} f(x; \mu) dx = 1$, $\forall \mu \in \mathbb{R}$.

Sin embargo, no todas las densidades que sean función de $x - \mu$ verifican la segunda ecuación. Recordemos que para estas familias:

$$\frac{\partial}{\partial x} f(x; \mu) = F'(x - \mu), \quad \frac{\partial}{\partial \mu} f(x; \mu) = -F'(x - \mu)$$

Entonces, la expresión (7.13) se escribe:

$$\begin{aligned}
& \int_{\mathbb{R}} \frac{\partial}{\partial x} f(x; \mu) x + f(x; \mu) - \frac{\partial}{\partial \mu} f(x; \mu) \mu dx = \\
& = 1 - \mu \int_{\mathbb{R}} (-F'(x - \mu)) dx - \int_{\mathbb{R}} (-F'(x - \mu)) x dx =
\end{aligned}$$

$$\begin{aligned}
&= 1 - \mu \frac{\partial}{\partial \mu} \left(\int_{\mathbb{R}} f(x; \mu) dx \right) - \frac{\partial}{\partial \mu} \left(\int_{\mathbb{R}} f(x; \mu) x dx \right) = \\
&= 1 - \frac{\partial}{\partial \mu} \left(\int_{\mathbb{R}} f(x; \mu) x dx \right)
\end{aligned}$$

donde se está suponiendo que existen las condiciones de regularidad necesarias para poder intercambiar $\frac{\partial}{\partial \mu}$ e integración sobre \mathbb{R} . Sin embargo, la expresión anterior no necesariamente es cero. Basta considerar un estimador equivariante que no tenga momento de primer orden. Como ejemplo podemos considerar un estimador cuya ley tuviera como densidad:

$$f(x; \mu) = \frac{1}{\pi} \frac{1}{1 + (x - \mu)^2}, \quad x \in \mathbb{R}$$

esto es, siguiera la distribución de Cauchy de parámetros $a = 1$ y $b = \mu$.

El ejemplo anterior hemos visto como el hecho de que la familia $N(\mu, 1)$ fuese estable conjuntamente frente a las transformaciones introducidas por V_1 y por V_2 permite restringirnos a una clase razonable de estimadores: aquellos cuya densidad respecto al volumen Riemanniano verifique la condición (7.8) para ambos campos. Hemos visto que las soluciones de estas ecuaciones no sólo determinan estimadores equivariantes sino que entre los equivariantes (esto es verifican (7.8) para V_1) para el grupo afín seleccionamos aquellos que además verifican (7.8) pero ahora tomando como campo V nuestro V_2 . Estamos siendo más exigentes que al requerir únicamente equivariancia frente a la acción del grupo afín.

Un problema diferente, y que se espera abordar en futuros estudios, es la resolución de la ecuación integro-diferencial (7.8) para diferentes familias paramétricas de distribuciones de probabilidad y diferentes campos vectoriales.

Con este ejemplo sobre la familia normal univariante $N(\mu, 1)$ damos por finalizado el capítulo y también esta memoria. Este capítulo debe verse como una breve presentación de las ideas de la estimación estable, unos criterios que creemos razonables a la hora de seleccionar un estimador de los parámetros de una familia. Esperamos que, en futuros trabajos, la conveniencia de los requerimientos que aquí se han presentado quede de manifiesto.

Capítulo 8

Conclusiones

8.1 Resumen de resultados de la memoria

En esta sección realizaremos un resumen de los principales resultados de este trabajo.

- En primer lugar, y vía un ejemplo concreto, se ha mostrado la conveniencia de seleccionar estimadores equivariantes, cuando la familia en estudio es invariante bajo la acción de algún grupo. Este requerimiento hace posible, aplicando criterios de mínimo riesgo, la selección de un estimador de mínimo riesgo entre los equivariantes, que consideraremos *óptimo*.
- Se ha demostrado que los elementos de una variedad \mathcal{C}^∞ pueden ser proyectados, según una discrepancia, localmente y de manera única sobre una subvariedad de la variedad dada. En general las discrepancias no tienen porque ser distancias o funciones de estas, pero en la mayoría de las situaciones que se han considerado sí que lo son.
- Se han proporcionado las definiciones de familia invariante en orden, pérdida invariante en orden y estimador invariante en orden frente a una familia de transformaciones en el espacio muestral.
- Se ha demostrado que la distancia de Rao es invariante en sentido débil de orden, como mínimo, el de la familia.
- El riesgo y el sesgo de un estimador invariante en orden están controlados, en orden, por cantidades que dependen del orden de la familia y del estimador.

- Se han reproducido, en orden y en la medida de lo posible, los resultados de Lehmann [27] de 1951.
- Por último, se ha introducido el concepto de estimación estable con el objetivo de establecer restricciones sobre la clase de estimadores a considerar.

8.2 Perspectivas de la Estimación Estable

En vista de los resultados obtenidos, parece conveniente exigir a un estimador que tenga orden de invariancia como mínimo el de la familia. Asimismo ocurre con el orden de estabilidad frente a la perturbación introducida por un campo. Cabe recordar que para la situación de invariancia clásica los órdenes de la familia, de la pérdida y del estimador relativos a la perturbación introducida por el campo $X(\theta) = \frac{d}{dt}g_t^*\theta|_{t=0}$, $\theta \in \Theta$, son todos $+\infty$.

Dentro de esta clase razonable de estimadores, en el sentido de estabilidad introducido en el capítulo 7, podemos imponer criterios de mínimo riesgo. Pero, tal y como se ha comentado anteriormente, antes de requerir condiciones que involucren a la función de pérdida, parece conveniente considerar la estabilidad o no de un estimador, relativa a las transformaciones introducidas por un campo.

Sin embargo, en este mismo capítulo también se pretende caracterizar la clase de estimadores que sean estables frente a una amplia clase de perturbaciones. En este sentido, cabe destacar la ecuación (7.8) y todo el proceso que ha llevado a su obtención. En futuros trabajos, se pretende estudiar esta ecuación para diferentes familias paramétricas de distribuciones de probabilidad y para distintas clases de perturbaciones, además de clarificar el papel que desempeñan las componentes normal y tangencial de un campo (donde estamos entendiendo normal y tangencial a Θ) en la estabilidad del estimador. Dentro de la amplia clase de perturbaciones a las que nos hemos referido anteriormente y además de los campos de Killing, posiblemente los campos de Killing en la subvariedad, los formados por uno de Killing en Θ y una contaminación en Δ y los campos que sean transportes paralelos de vectores constantes a lo largo de una geodésica de Θ aporten resultados destacados al problema de la estimación estable.

Los campos de Killing en la subvariedad, localmente definen isometrías. Esta condición hace esperar que sus transformaciones asociadas sean una clase de perturbaciones del modelo con buenas propiedades. Sin embargo, pueden haber otros campos que verifiquen esta condición. Más aún: pueden existir campos para los que la transformación obtenida sea una isometría *en orden*,

concepto que no ha sido introducido en esta memoria pero que podemos enmarcar claramente en la filosofía del trabajo.

En definitiva, una vez que se ha seleccionado una clase de estimadores razonables y escogido una pérdida natural para el problema en estudio, podemos exigir otras propiedades de optimalidad (riesgo, sesgo, comportamiento asintótico...). De esta manera, se obtiene un proceso de selección de estimadores paralelo al que se acepta para la situación de invariancia clásica y que fue comentado en la introducción de este trabajo.

Tal y como se han formulado los resultados de esta memoria, en algunos puntos existe aún dependencia de la variedad ambiente, por ejemplo la condición de ser de campo de Killing depende de la variedad en la que se considera. Sería deseable, a nivel práctico, centrarse únicamente en la subvariedad, esto es, en el modelo paramétrico de partida.

Cuestiones interesantes pueden aparecer cuando nos preguntemos sobre el comportamiento de un estimador estable frente a una determinada perturbación si consideramos otra perturbación distinta. En el caso concreto de los equivariantes, sería interesante caracterizar la familia de perturbaciones (diferentes de la transformación inducida por invariancia) frente a la que éstos serían estables o estables en orden. Más aún, estudiar qué condiciones deberían cumplir para que sean éstos sean estables, en orden o no, frente a cualquier perturbación.

Finalmente, en familias paramétricas de distribuciones de probabilidad para las que no existan grupos que las dejen invariantes, se espera que todas estas ideas aporten resultados destacados al problema de la estimación de los parámetros del modelo. Hemos introducido el vocabulario y los conceptos, esto es, el marco necesario para modelizar estas situaciones. Ahora podremos concretar estas ideas en las familias que lo requieran.

Apéndice A

Algunos aspectos de Geometría Diferencial

Este apéndice está dedicado a introducir gran parte de las definiciones relevantes del trabajo. Aunque la mayoría de las definiciones ya son conocidas para variedades modeladas en \mathbb{R}^n , es agradable verificar que se pueden recuperar en el caso de variedades modeladas en un espacio de Banach no necesariamente de dimensión finita. Observaremos en este capítulo que básicamente las variedades deben ser modeladas en un espacio en el que podamos realizar Cálculo Diferencial.

A.1 Variedades

Los espacios de parámetros que consideraremos en este trabajo se supondrán con estructura de variedad. Una variedad puede entenderse como un espacio que localmente es de Banach. Ejemplos triviales de variedad son el de un espacio Euclídeo n -dimensional o el de un espacio de Hilbert. La definición formal de variedad es la siguiente.

Definición A.1.1. Sea Δ un conjunto. Un *atlas de clase \mathcal{C}^k* ($k \geq 0$) en Δ es una colección de pares $\{(V_i, y_i)\}_{i \in I}$ (I un conjunto conveniente de índices) de manera que:

1. $\{V_i\}_{i \in I}$ es un recubrimiento de Δ .
2. Cada y_i es una biyección de V_i en un subconjunto abierto $y_i(V_i)$ de un espacio de Banach \mathbf{E}_i . Además para cualquier par de índices i, j , $y_i(V_i \cap V_j)$ es un conjunto abierto de \mathbf{E}_i .

3. $\forall i, j$, la aplicación:

$$y_j \circ y_i^{-1} : y_i(V_i \cap V_j) \longrightarrow y_j(V_i \cap V_j)$$

es de clase \mathcal{C}^k .

Observemos que en la condición 2 de la definición de atlas no se requiere que los espacios \mathbf{E}_i sean el mismo para todo i . Ni siquiera se les pide que sean topológicamente isomorfos. Sin embargo, es posible demostrar que si $k \geq 1$, basta restringirnos al caso en que todos los \mathbf{E}_i son un mismo espacio de Banach \mathbf{E} . Más detalles al respecto se pueden encontrar en Lang [26].

Definición A.1.2. Cada par (V_i, y_i) es una *carta* del atlas. Además si $p \in V_i$, diremos que (V_i, y_i) es una *carta en* p .

Definición A.1.3. Una carta (V, y) es \mathcal{C}^k -compatible con el atlas de clase \mathcal{C}^k $\{(V_i, y_i)\}_{i \in I}$, si cada aplicación $y_i \circ y^{-1}$ (en aquella intersección en la que esté definida) es de clase \mathcal{C}^k .

Definición A.1.4. Diremos que dos atlas de clase \mathcal{C}^k son *compatibles* si cada carta de uno es compatible con el otro atlas.

Observemos que la relación de compatibilidad entre atlas es una relación de equivalencia. De esta manera, tiene sentido introducir la siguiente definición.

Definición A.1.5. Diremos que una clase de equivalencia de la relación de compatibilidad entre atlas \mathcal{C}^k de Δ define una estructura de \mathcal{C}^k -variedad o de variedad de clase \mathcal{C}^k en Δ . Si todos los espacios \mathbf{E}_i son el mismo (que notaremos \mathbf{E}) diremos que la \mathcal{C}^k -variedad *está modelada en* \mathbf{E} o que Δ es una \mathbf{E} -variedad \mathcal{C}^k .

Sea Δ una \mathcal{C}^k -variedad y $\{(V_i, y_i)\}_{i \in I}$ representante de la clase de equivalencia entre atlas que define a la variedad. Adoptaremos como convenio, en las definiciones y explicaciones que siguen, que una carta siempre será una carta de este atlas.

Definición A.1.6. Diremos que un subconjunto $A \subset \Delta$ es un *abierto* de Δ si y sólo si $y(A \cap V)$ es abierto para toda carta (V, y) .

Con esta definición de abierto, se dota a Δ de una topología en la que para cada carta (V, y) , V es abierto en Δ y las aplicaciones y son homeomorfismos. Podríamos además estar interesados en la suavidad de estas aplicaciones. Esto nos lleva a definir lo que entenderemos por una aplicación de clase \mathcal{C}^r entre dos variedades cualquiera.

Definición A.1.7. Sean Δ_1 y Δ_2 dos variedades de clase k_1 y k_2 respectivamente. Sea $A \subset \Delta_1$ un conjunto abierto y $f : A \rightarrow \Delta_2$. Diremos que f es una aplicación de *clase* \mathcal{C}^r , $r \leq \min\{k_1, k_2\}$, en $p \in A$ si existe una carta (U, ϕ) en p y una carta (V, ψ) en $f(p)$ tales que $f(A \cap U) \subset V$, y la aplicación:

$$\psi \circ f \circ \phi^{-1} : \phi(A \cap U) \rightarrow \psi(V)$$

sea de clase \mathcal{C}^r . f es de *clase* \mathcal{C}^r en A si lo es en todo punto de A y cuando $A = \Delta_1$ únicamente diremos que f es de *clase* \mathcal{C}^r .

Es posible verificar que si esta condición es cierta, también lo es para cualquier elección de cartas (U, ϕ) en p y (V, ψ) en $f(p)$ tales que $f(U) \subset V$.

Finalmente comentar que únicamente vamos a estar interesados en variedades de clase \mathcal{C}^∞ y para las que la topología que hemos introducido sea Hausdorff. Entenderemos entonces que una variedad es un objeto verificando las condiciones de la definición (A.1.5) con $k = \infty$ y para el que la topología introducida en (A.1.6) sea Hausdorff.

A.2 Vectores tangentes. Espacio tangente.

El objetivo de esta sección es introducir los elementos necesarios para poder generalizar, en la medida de lo posible, la derivada direccional ordinaria de \mathbb{R}^n .

Sea Δ una variedad y $p \in \Delta$. Consideremos tripletas de la forma (U, ϕ, x) , siendo (U, ϕ) una carta en p y x un elemento del espacio de Banach donde $\phi(U)$ se encuentra.

Definición A.2.1. Diremos que (U, ϕ, x) y (V, ψ, y) están *relacionados* si y sólo si:

$$(\psi \circ \phi^{-1})|_{\phi(p)}(x) = y$$

Observemos que esta relación es de equivalencia en virtud de la regla de la cadena.

Definición A.2.2. Una clase de equivalencia de esta relación es un *vector tangente* en p .

Definición A.2.3. El conjunto de todos los vectores tangentes en p es el *espacio tangente* en p y lo notaremos como $T_p(\Delta)$.

Sea ahora $p \in \Delta$ y (U, ϕ) una carta en p . Veremos que, fijada esta carta, es posible establecer una biyección de $T_p(\Delta)$ en un espacio de Banach.

Recordemos que para variedades de clase $k \geq 1$ podemos suponer que todos los espacios \mathbf{E}_i son el mismo y notarlo \mathbf{E} . Por tanto, el espacio de Banach que estará en biyección con $T_p(\Delta)$ es \mathbf{E} , $\forall p \in \Delta$. Además, si (V, ψ, y) es un elemento de la clase de equivalencia de $[(V, \psi, y)]$, existe $x \in E$ tal que (V, ψ, y) y (U, ϕ, x) están relacionadas. En efecto, basta tomar $x = (\psi \circ \phi^{-1})'_{|\phi(p)}(y)$. Podremos así escribir cada una de las clases de equivalencia de $T_p(\Delta)$ como $[(U, \phi, x)]$ mientras que la biyección entre $T_p(\Delta)$ y \mathbf{E} , que notaremos J_p , queda:

$$\begin{aligned} J_p : \quad T_p(\Delta) &\longrightarrow \mathbf{E} \\ [(U, \phi, x)] &\longmapsto x \end{aligned}$$

Notación A.2.1. Como convenio consideraremos que con (U, ϕ, x) estamos indicando la clase de equivalencia $[(U, \phi, x)]$. Pero habitualmente y por simplicidad de notación, escribiremos X_p en lugar de (U, ϕ, x) . Finalmente diremos que x es el representante de X_p para la carta (U, ϕ) .

Observemos que por medio de esta biyección es posible dotar a $T_p(\Delta)$ estructura de espacio vectorial topológico, demostrándose que esta estructura es independiente de la carta (U, ϕ) seleccionada.

A.3 Aplicación Jacobiana. Hessiano.

Es bien sabido que para una aplicación suficientemente suave entre espacios de Banach podemos considerar su derivada. En el caso de disponer de una aplicación entre variedades podemos introducir la siguiente definición.

Definición A.3.1. Sea $f : \Delta_1 \longrightarrow \Delta_2$ una aplicación entre variedades de clase \mathcal{C}^r , $r \geq 1$. El *jacobiano* de f en el punto $p \in \Delta_1$, que notaremos $(f_*)_p$ es la única aplicación de $T_p(\Delta_1)$ en $T_{f(p)}(\Delta_2)$ cumpliendo que si (U, ϕ) es una carta en p , (V, ψ) es una carta en $f(p)$ tal que $f(U) \subset V$ y X es el vector tangente en p representado por x en la carta (U, ϕ) , entonces $(f_*)_p(X)$ es el vector tangente en $f(p)$ representado por $(\psi \circ f \circ \phi^{-1})'_{|\phi(p)}(x)$.

Escribiremos f_* para la aplicación jacobiana de f si no es necesario indicar el punto en el que se está considerando.

Observemos que la aplicación anterior está bien definida gracias a la estructura de espacio vectorial topológico con la que hemos dotado tanto a $T_p(\Delta_1)$ como a $T_{f(p)}(\Delta_2)$.

El elemento $(\psi \circ f \circ \phi^{-1})'_{|\phi(p)}(x)$ no es nada más que la derivada direccional en la dirección x de la función $(\psi \circ f \circ \phi^{-1})$ evaluada en el punto $\phi(x)$. Si \mathbf{E} es \mathbb{R}^n sabemos que cuando la derivada direccional existe en cada dirección $x \in \mathbb{R}^n$, también existen las derivadas parciales de la función, esto es las derivadas direccionales en la dirección de los vectores de la base canónica e_1, \dots, e_n . Si $f : \Delta_1 \longrightarrow \Delta_2$ es una aplicación \mathcal{C}^∞ entre variedades y Δ_1 está modelada sobre \mathbb{R}^n podremos considerar entonces las derivadas parciales:

$$D_1(\psi \circ f \circ \phi^{-1})_{|\phi(p)}, \dots, D_n(\psi \circ f \circ \phi^{-1})_{|\phi(p)}$$

que escribiremos:

$$f_*\left(\frac{\partial}{\partial \phi^1}(p)\right), \dots, f_*\left(\frac{\partial}{\partial \phi^n}(p)\right)$$

De esta manera, hemos recuperado expresiones ya conocidas para el jacobiano de una aplicación entre variedades finito-dimensionales.

El siguiente punto a considerar es del Hessiano de una aplicación a valores reales. Antes necesitamos introducir una serie de definiciones previas.

Sea Δ una variedad y $f : \Delta \longrightarrow \mathbb{R}$ de clase \mathcal{C}^r , $r \geq 1$. Sea $p \in \Delta$ y (U, ϕ) una carta en p . Siguiendo la definición anterior, observemos que $\forall X = (U, \phi, x) \in T_p(\Delta)$, $f_{*,p}(X)$ es el vector tangente de $T_{f(p)}(\mathbb{R})$ que se identifica por medio de la carta identidad con:

$$(f \circ \phi^{-1})'_{|\phi(p)}(x)$$

Esta observación, nos lleva a introducir:

Definición A.3.2. La *acción del vector* $X \in T_p(\Delta)$ sobre f , que notaremos Xf , como $Xf := (f \circ \phi^{-1})'_{|\phi(p)}(x)$.

Definición A.3.3. Diremos que X es un *campo vectorial de dominio* $A \subset \Delta$ si y sólo si a cada $p \in A$, X le asigna un elemento de $T_p(\Delta)$ que denotaremos $X(p)$ o bien X_p .

Diremos que el campo X es \mathcal{C}^∞ en A si y sólo si A es abierto y para cada función f a valores reales de clase \mathcal{C}^∞ en B , un conjunto abierto de Δ , la asignación $p \mapsto X_p(f)$ es de clase \mathcal{C}^∞ en $A \cap B$.

Podemos ahora introducir el objeto que va a desempeñar el papel del Hessiano en un espacio de Banach.

Definición A.3.4. Sea f una función \mathcal{C}^r , $r \geq 1$ a valores reales y de dominio un entorno abierto A de un punto $p \in \Delta$. El *Hessiano de f en el punto $p \in \Delta$* es la aplicación:

$$\begin{aligned} Hess^{f,p} : T_p(\Delta) \times T_p(\Delta) &\longrightarrow \mathbb{R} \\ (X_p, Y_p) &\longmapsto Y_p(Xf) \end{aligned}$$

donde X es un campo tangente de dominio $A \ni p$ cuyo valor en p es X_p .

Esta aplicación va a desempeñar un papel importante tal y como veremos en el capítulo 2.

A.4 Conexiones

En la sección anterior hemos definido lo que entendemos por la acción de un elemento del espacio tangente X sobre una función $f : \Delta \longrightarrow \mathbb{R}$. Ya hemos comentado que si repasamos la definición para $\Delta = \mathbb{R}^n$ vemos que Xf es en realidad la derivada de la función f en la dirección $x \in \mathbb{R}^n$.

En esta sección vamos a analizar la variación de un campo vectorial X en la dirección de otro campo Y .

Sea Δ una variedad \mathcal{C}^∞ y X e Y dos campos vectoriales \mathcal{C}^∞ de dominio A .

Definición A.4.1. Una *conexión* en Δ es un operador, que notaremos D que asigna al par (X, Y) un nuevo campo vectorial $D_X Y$ de dominio A de manera que si Z es otro campo \mathcal{C}^∞ en A y f es una función a valores reales en A , D cumple las siguientes propiedades:

1. $D_X(Y + Z) = D_X Y + D_X Z$
2. $D_{X+Y} Z = D_X Z + D_Y Z$
3. $D_{fX} Y = f D_X Y$
4. $D_X(fY) = (Xf)Y + f(D_X Y)$

Definición A.4.2. Diremos que $D_X Y$ es la *derivada covariante* de Y respecto a X .

La pregunta que nos podemos formular es cuándo existe conexión en una variedad \mathcal{C}^∞ dada. En la sección (A.7) recordaremos que en variedades Riemannianas existe una conexión *natural* en un sentido que allí especificaremos. Para una variedad arbitraria deberemos suponer simplemente que existe. El concepto de conexión va a permitirnos introducir el de paralelismo a lo largo de una curva y el de geodésicas en una variedad \mathcal{C}^∞ .

Definición A.4.3. Una curva γ en Δ es una aplicación de clase \mathcal{C}^∞ de un intervalo abierto $(a, b) \subset \mathbb{R}$ en Δ .

Definición A.4.4. Diremos que una aplicación $\gamma : [a, b] \subset \mathbb{R} \rightarrow \Delta$ es un *segmento de curva* si y sólo si existe una curva $\bar{\gamma} : (\alpha, \beta) \rightarrow \Delta$ de manera que $\bar{\gamma}|_{(a,b)} \equiv \gamma$.

Como convenio, en este trabajo vamos a suponer que todas las curvas que consideramos son segmentos de curva. De esta manera, los vectores tangentes en los extremos estarán siempre bien definidos.

Sea γ una curva. Por ser de clase \mathcal{C}^∞ , para todo $t \in (a, b)$ podemos considerar $\gamma_*(\frac{d}{dt})$ siendo γ_* la aplicación jacobiana de γ y $\frac{d}{dt}$ la derivación estándar de \mathbb{R} .

Definición A.4.5. El vector *tangente a la curva γ en t* es $\gamma_*(\frac{d}{dt})$, que notaremos $T(t)$ o $T_\gamma(t)$ si fuera necesario.

Definición A.4.6. Diremos que un campo vectorial Y es \mathcal{C}^∞ sobre una curva γ si y sólo si la asignación $t \mapsto Y(t) \in T_{\gamma(t)}(\Delta)$ es de clase \mathcal{C}^∞ .

Definición A.4.7. Diremos que un campo Y que sea \mathcal{C}^∞ a lo largo de una curva γ es *paralelo a lo largo de γ* si y sólo si $D_T Y = 0$ en γ .

Definición A.4.8. Diremos que γ es una *geodésica* si y sólo si la derivada covariante $D_T T = 0$ en γ .

Observemos que las geodésicas generalizan el concepto de línea recta en el espacio Euclídeo pues el vector tangente a un segmento rectilíneo es paralelo a lo largo de la curva.

El siguiente teorema prueba la existencia local de geodésicas en una variedad con conexión y puede encontrarse demostrado en cualquier texto básico de Geometría Diferencial.

Teorema A.4.1. Sea $p \in \Delta$ y $X \in T_p(\Delta)$. Entonces, para cualquier $\alpha \in \mathbb{R}$, $\exists r > 0$ y una única curva γ_X definida en $[\alpha - r, \alpha + r]$ de manera que $\gamma_X(\alpha) = p$, $T_{\gamma_X}(\alpha) = X$ y γ_X es una geodésica.

En la sección (A.7) de este capítulo veremos, una vez que el concepto de longitud pueda ser introducido veremos que además las geodésicas son localmente las curvas de longitud más corta que unen dos puntos.

A.5 Aplicación exponencial.

Sea Δ una variedad con conexión. Si revisamos el teorema A.4.1 veremos que para cada vector X tangente a Δ en p existe una única geodésica γ_X de la conexión D que está definida en un entorno de $0 \in \mathbb{R}$ de manera que $\gamma_X(0) = p$ y tiene vector tangente X en el origen. Debido a la naturaleza de las ecuaciones que determinan las geodésicas en una variedad con conexión, se tiene que $g_X(1)$ está bien definida si X está suficientemente cerca del origen. Sea entonces \mathcal{D}_p el conjunto de vectores $X \in T_p(\Delta)$ para los que la geodésica γ_X está definida como mínimo en $[0, 1]$.

Definición A.5.1. La *aplicación exponencial* en $p \in \Delta$ es la función:

$$\begin{aligned} \exp_p : \mathcal{D}_p &\longrightarrow \Delta \\ Y &\longmapsto \gamma_Y(1) \end{aligned}$$

Para más propiedades y resultados sobre la aplicación exponencial debemos esperar a la sección correspondiente a variedades Riemannianas o bien consultar cualquier texto básico de Geometría Diferencial.

A.6 Subvariedades

Cabe plantearse si la estructura de variedad de Δ induce alguna sobre sus subconjuntos. Vamos a ver que, en efecto, así es.

Definición A.6.1. Sean Δ_1 y Δ_2 dos variedades. Una *inmersión* es una aplicación $f : \Delta_1 \longrightarrow \Delta_2$ de clase \mathcal{C}^∞ de manera que $(f_*)_p$ sea 1 – 1 para todo $p \in \Delta_1$.

Definición A.6.2. Sean Δ y Θ dos variedades y supongamos que $\Theta \subset \Delta$. Diremos que Θ es una *subvariedad* de Δ si y sólo si:

1. Θ es un subespacio topológico de Δ .
2. La inclusión $i : \Theta \hookrightarrow \Delta$ es una inmersión.

En ocasiones se consideran subvariedades para las que la inclusión es una inmersión global. Pasamos a enunciar este concepto, aunque en este trabajo no haremos uso directo de él.

Definición A.6.3. Siguiendo con las notaciones anteriores, diremos que una inmersión $f : \Delta_1 \longrightarrow \Delta_2$ es *global* si y sólo si:

1. f es una inmersión $1 - 1$.
2. La aplicación inducida $\Delta_1 \rightarrow f(\Delta_1)$ es un homeomorfismo en el subespacio topológico $f(\Delta_1) \subset \Delta_2$.

Si Θ es una subvariedad de Δ para la que la inclusión $\Theta \hookrightarrow \Delta$ es una inmersión global diremos que Θ es una *subvariedad globalmente inmersa en Δ* .

Para ilustrar la definición de subvariedad, consideremos el siguiente ejemplo.

Ejemplo A.6.1. Sea Θ el subconjunto de \mathbb{R}^2 dado por la figura A.1. Observemos que la inclusión $i : \Theta \hookrightarrow \mathbb{R}^2$ es una inmersión. Sin embargo, Θ no es un subespacio topológico de \mathbb{R}^2 pues cualquier abierto de Θ conteniendo a θ no puede ser escrito como la intersección de Θ y de un conjunto abierto de \mathbb{R}^2 .



Figura A.1: Subconjunto inmerso en \mathbb{R}^2 que no es un subespacio topológico de \mathbb{R}^2 .

Los conceptos de subvariedad, subvariedad globalmente inmersa, inmersiones e inmersiones globales se encuentran discutidos con detalle en las páginas 19 a 21 de la referencia O'Neil [32].

A.7 Variedades Riemannianas

Ya hemos visto, en la sección (A.1), que es posible definir una topología en una variedad Δ a través de las cartas locales. Podemos preguntarnos si existen variedades que también tengan estructura de espacio métrico y más aún si la topología definida por la distancia y la de la variedad coinciden. En esta sección, que está basada en la parte correspondiente también a variedades Riemannianas de Lang [26], veremos que todas estas preguntas tienen respuesta afirmativa.

Sea Δ una variedad no necesariamente de dimensión finita.

Definición A.7.1. Una *métrica Riemanniana* g en Δ es una aplicación que a todo punto $p \in M$ le asocia un producto escalar $g_p : T_p(M) \times T_p(M) \rightarrow \mathbb{R}$ de manera que cumpla la siguiente propiedad de diferenciabilidad: si X e Y son campos \mathcal{C}^∞ de dominio A , entonces:

$$p \mapsto g_p(X_p, Y_p)$$

es una aplicación de clase \mathcal{C}^∞ en A .

Definición A.7.2. Una *variedad Riemanniana* consiste en una variedad Δ equipada con una métrica Riemanniana en Δ .

Las variedades Riemannianas van a desempeñar un papel importante en este trabajo pues, bajo ciertas condiciones, una familia paramétrica de distribuciones de probabilidad tiene estructura de variedad Riemanniana dada por su métrica informativa. Más detalles al respecto pueden ser consultados en Amari [3], Atkinson y Mitchel [5] o en Burbea y Rao [11].

Vamos a ver ahora que es posible definir una estructura de espacio métrico en un Δ a través de la métrica Riemanniana. Sea $p \in M$ y X un campo vectorial de dominio $A \ni p$ cuyo valor en p es X_p .

Definición A.7.3. Definimos la *norma* de X_p como $\|X_p\| = \sqrt{g_p(X_p, X_p)}$

El siguiente paso es definir la longitud de una curva, esto es de una aplicación γ de clase \mathcal{C}^∞ de un intervalo abierto $(a, b) \subset \mathbb{R}$ en Δ . Sea $T_\gamma(t) = \gamma_*\left(\frac{d}{dt}\right)$ el vector tangente a la curva γ en t . Introducimos:

Definición A.7.4. La *longitud de la curva* γ desde a hasta b , que notaremos $\|\gamma\|_a^b$, es:

$$\|\gamma\|_a^b = \int_a^b \sqrt{g_{\gamma(t)}(T(t), T(t))} dt$$

Si γ fuera una curva \mathcal{C}^∞ a trozos, definiríamos la longitud de γ como la suma finita de las longitudes de sus trozos \mathcal{C}^∞ . Se comprueba que $\|\gamma\|_a^b$ no depende de la parametrización escogida.

Supongamos que Δ sea una variedad conexa. Introducimos la aplicación:

$$\begin{aligned} \rho : \Delta \times \Delta &\longrightarrow [0, +\infty) \\ (p, q) &\longmapsto \inf \{ \|\gamma\| \} \end{aligned}$$

donde γ es una curva \mathcal{C}^∞ de p a q . Se verifica que ρ define una pseudo-distancia en Δ , esto es $\forall p, q, r \in \Delta$ la aplicación ρ cumple:

1. $\rho(p, q) \geq 0$ y $\rho(p, p) = 0$.
2. $\rho(p, q) = \rho(q, p)$.
3. $\rho(p, q) \leq \rho(p, r) + \rho(r, q)$.

Estamos ahora en posición de analizar la relación entre las topologías propia de la variedad y la inducida por ρ . Tal y como pueden encontrarse probados en la página 70 de Hicks [23] en la 186 de Lang [26] o en la 115 de Petersen [33]:

Teorema A.7.1. *La topología obtenida a través de la pseudo-distancia en una variedad Riemanniana es la misma que la topología de la variedad.*

Finalmente debemos averiguar bajo qué condiciones la pseudo-distancia ρ es una distancia. Observemos que la condición que falta por verificarse es que $\rho(p, q) = 0$ implique $p = q$.

Proposición A.7.2. *En una variedad Riemanniana conexa Δ la pseudo-distancia ρ es una distancia.*

Si repasamos la demostración de este último resultado vemos que una condición esencial en la demostración es que la topología de la variedad sea Hausdorff. Recordemos que este requerimiento ha quedado incorporado en la definición de variedad tal y como se vio en la sección (A.1). Finalmente, como convenio entenderemos que una variedad Riemanniana siempre es conexa y que por tanto la pseudo-distancia ρ siempre es una distancia.

En una variedad Riemanniana vamos a estar interesados en una conexión destacada o natural en el sentido de la siguiente definición.

Definición A.7.5. Una *conexión Riemanniana* en una variedad Riemanniana Δ es una conexión D cumpliendo:

1. $D_X Y - D_Y X = [X, Y]$.
2. $Z \cdot \langle X, Y \rangle = \langle D_Z X, Y \rangle + \langle X, D_Z Y \rangle$

para campos vectoriales X, Y y Z de dominio común.

Se demuestra que en una variedad Riemanniana existe una única conexión Riemanniana. Habitualmente nos referiremos a esta conexión como *Conexión de Levi-Civita*.

Hemos anunciado anteriormente que las geodésicas en una variedad Riemanniana son localmente las curvas de longitud mínima que unen dos puntos. Este conocido resultado se basa en:

Definición A.7.6. Sea Δ una variedad con conexión afín y p un punto de Δ . Diremos que un entorno abierto N_0 del origen de $T_p(\Delta)$ es *normal* si y sólo si:

1. La aplicación exponencial, \exp , es un difeomorfismo de N_0 en un entorno abierto N_p de p en Δ .
2. Si $X \in N_0$ y $0 \leq t \leq 1$, entonces $tX \in N_0$

Definición A.7.7. Diremos que un entorno N_p de p en Δ es un *entorno normal* de p si y sólo si $N_p = \exp(N_0)$ siendo N_0 un entorno normal de $0 \in T_p(\Delta)$.

Siempre que nos situemos en entornos normales de un punto las geodésicas serán las curvas de longitud mínima que unan dos punto cualquiera en este entorno. El siguiente teorema así lo afirma pero además garantiza la existencia de normales de cualquier punto de una variedad con conexión afín.

Lema A.7.3. Sea (Δ, ρ) una variedad Riemanniana. Para cada $p \in \Delta$ existe $r(p) > 0$ tal que si $0 < r < r(p)$, $B_r(p)$ cumple las propiedades:

1. $B_r(p)$ es un entorno normal de cada uno de sus puntos.
2. Sean $a, b \in B_r(p)$ y γ_{ab} la única geodésica en $B_r(p)$ uniendo a y b. Entonces γ_{ab} es el único segmento de curva en Δ de longitud $\rho(a, b)$ que une a y b.

La demostración de este enunciado en concreto puede encontrarse en la página 53 de la referencia Helgason [22].

Una vez establecida la estructura métrica de una variedad Riemanniana (Δ, ρ) , debemos considerar si las subvariedades de Δ pueden heredar de alguna forma esta estructura.

Sea g la métrica Riemanniana en Δ en el sentido de la definición A.7.1. Si Θ es una subvariedad de Δ , $\forall \theta \in \Theta$ cada espacio tangente $T_\theta(\Theta)$ puede ser considerado como un subespacio de $T_\theta(\Delta)$. De esta manera podemos considerar una aplicación que claramente es bilineal, simétrica:

$$\begin{aligned} \bar{g}_\theta : T_\theta(\Theta) \times T_\theta(\Theta) &\longrightarrow [0, +\infty) \\ (X_\theta, Y_\theta) &\longmapsto g_p(X_\theta, Y_\theta) \end{aligned}$$

Definición A.7.8. Si la aplicación anterior \bar{g}_θ es una métrica Riemanniana en Θ en el sentido de la definición A.7.1 diremos que Θ es una *subvariedad Riemanniana* de Δ .

Para esta métrica que acabamos de introducir también dispondremos de una conexión Riemanniana que llamaremos inducida y notaremos \overline{D} y que en general no coincidirá con la de la variedad. La diferencia entre una y otra queda evaluada de la siguiente manera.

Sean X e Y campos vectoriales de clase \mathcal{C}^∞ de dominio $A \in \Theta$ y tangentes a Θ . Introducimos:

$$V(X, Y) = D_X Y - \overline{D}_X Y$$

donde D es la conexión Riemanniana en Δ dada por g y \overline{D} es la conexión Riemanniana en Θ dada por \overline{g} . Se demuestra que V es una aplicación bilineal y simétrica que obviamente toma valores vectoriales y para la que la asignación:

$$A \ni \theta \mapsto V(X_\theta, Y_\theta) \quad (\text{A.1})$$

es de clase \mathcal{C}^∞ donde X_θ e Y_θ son los valores de los campos X e Y en θ . La ecuación (A.1) es la *ecuación de Gauss* y la aplicación V es *el segundo tensor fundamental*.

A.8 Difeomorfismos. Isometrías.

Una vez definida la noción de distancia en una variedad, podemos considerar las aplicaciones entre variedades Riemannianas que conserven la distancia. Estas son las isometrías.

Definición A.8.1. Sean Δ_1 y Δ_2 dos variedades. Un *difeomorfismo* es una aplicación $f : \Delta_1 \longrightarrow \Delta_2$ que sea 1-1 y de manera que f y f^{-1} son \mathcal{C}^∞ .

Definición A.8.2. Sean (Δ_1, ρ_g) y (Δ_2, ρ_h) dos variedades Riemannianas y $\Phi : M \longrightarrow N$ un difeomorfismo. Sea $(\Phi_*)_p$ la aplicación jacobiana de Φ en el punto $p \in \Delta_1$ y supongamos que se cumple:

$$h_{\Phi(p)}((\Phi_*)_p(X_p), (\Phi_*)_p(Y_p)) = g_p(X_p, Y_p), \quad \forall X_p, Y_p \in M_p, \quad \forall p \in M$$

Diremos, en tal caso, que Φ es una *isometría* entre Δ y Δ_2 .

Observemos que si Φ es una isometría, también lo es Φ^{-1} . Además, si Φ es una isometría, es inmediato demostrar que preserva distancias.

A.9 Campos de Jacobi

Con las mismas notaciones de las secciones previas, sea Δ una variedad Riemanniana y ∇ la conexión de Levi-Civita en Δ . Sea $p \in \Delta$ y Q el conjunto de los $(s, t) \in \mathbb{R}^2$ tales que \exp_p está definida en $s \zeta(t)$ para cualquier curva $t \mapsto \zeta(t)$ de $T_p(\Delta)$. En toda esta sección, si no se menciona de forma explícita lo contrario, el par (s, t) será siempre un elemento de Q .

Consideremos la aplicación $c_p : Q \mapsto \Delta$ dada por:

$$c_p(s, t) = \exp_p(s \zeta(t))$$

Revisando la definición de la aplicación exponencial, se verifica que $\{c_p(\cdot, t)\}$ es una familia uniparamétrica de geodésicas en Δ , es decir, para cada valor de t , $c_p(s, t)$, es una geodésica respecto a la conexión ∇ de Δ . Introducimos:

$$c'_p = (c_p)_*\left(\frac{\partial}{\partial s}\right), \quad \dot{c}_p = (c_p)_*\left(\frac{\partial}{\partial t}\right),$$

donde $\frac{\partial}{\partial s}$ y $\frac{\partial}{\partial t}$ son los operadores de derivación parcial ordinarios en \mathbb{R}^2 . Observemos que c'_p es un campo vectorial tangente a lo largo de las curvas geodésicas $c_p(\cdot, t)$ mientras que \dot{c}_p es tangente a las curvas $c_p(s, \cdot)$. Estudiemos con más detalle los campos \dot{c}_p y c'_p . Se tiene que:

$$\frac{\partial}{\partial s} \langle c'_p, \dot{c}_p \rangle = \langle \frac{\nabla}{\partial s} c'_p, \dot{c}_p \rangle + \langle c'_p, \frac{\nabla}{\partial s} \dot{c}_p \rangle,$$

donde $\frac{\nabla}{ds}$ es la derivada covariante a lo largo de $c(\cdot, t)$. Por otra parte, fijado t la aplicación $s \mapsto c_p(s, t)$ es una geodésica y por tanto $\frac{\nabla}{\partial s} c'_p = 0$. Además la conexión de Levi-Civita es de torsión nula y el paréntesis de Lie $[c'_p, \dot{c}_p] = 0$. Recordemos que $[c'_p, \dot{c}_p]f = (\frac{\partial}{\partial s} \frac{\partial}{\partial t} - \frac{\partial}{\partial t} \frac{\partial}{\partial s})f$ para toda función f de clase \mathcal{C}^2 sobre M . De esta manera, los campos vectoriales c'_p y \dot{c}_p cumplen:

$$\frac{\nabla}{dt} c'_p = \frac{\nabla}{ds} \dot{c}_p \tag{A.2}$$

siendo ahora $\frac{\nabla}{dt}$ es la derivada covariante a lo largo de $t \mapsto c_p(s, t)$.

Antes de continuar este estudio sobre c'_p y \dot{c}_p , recordemos la definición del tensor de curvatura asociado a ∇ .

Definición A.9.1. Sea Δ una variedad Riemanniana y ∇ la conexión de Levi-Civita en Δ . Dados X, Y y Z campos vectoriales, definimos el *tensor de curvatura de ∇* , R , según:

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y, X]} Z$$

Sea ahora X un campo vectorial definido a lo largo de la geodésica $s \mapsto c_p(s, t)$. Por ser $[c'_p, \dot{c}_p] = 0$, se tiene:

$$R(c'_p, \dot{c}_p)X = \frac{\nabla}{dt} \frac{\nabla}{ds} X - \frac{\nabla}{ds} \frac{\nabla}{dt} X$$

Tomando en la expresión anterior como X el campo c'_p y debido a que, fijado t , la aplicación $s \mapsto c_p(s, t)$ es una geodésica, obtenemos:

$$0 = \frac{\nabla}{ds} c'_p = \frac{\nabla}{dt} \frac{\nabla}{ds} c'_p = \frac{\nabla}{ds} \frac{\nabla}{dt} c'_p + R(c'_p, \dot{c}_p) c'_p$$

De aquí, y debido a (A.2), el campo vectorial \dot{c}_p a lo largo de la geodésica $s \mapsto c_p(s, t)$ verifica la ecuación diferencial de segundo orden:

$$\frac{\nabla^2}{ds^2} \dot{c}_p + R(c'_p, \dot{c}_p) c'_p = 0. \quad (\text{A.3})$$

Esta ecuación nos lleva a introducir el siguiente concepto.

Definición A.9.2. Sea $c = c(s)$ una geodésica en Δ e Y un campo vectorial a lo largo de c . Diremos que Y es *un campo de Jacobi a lo largo de c* si verifica la ecuación:

$$\frac{\nabla^2}{ds^2} Y + R(c', Y) c' = 0$$

donde R es el tensor de curvatura, c' denota el campo vectorial tangente a c y $\frac{\nabla}{ds}$ es la derivada covariante a lo largo de c . Esta ecuación se conoce como *ecuación de Jacobi*.

Observemos que con la ecuación de Jacobi se describe la relación entre la curvatura y el comportamiento de las curvas geodésicas en un entorno de una geodésica dada.

La ecuación (A.3) informa de que el campo c'_p es un campo de Jacobi a lo largo de la geodésica $s \mapsto c_p(s, t)$. Este es un resultado destacado que se ha utilizado en este trabajo en repetidas ocasiones.

Para variedades Riemannianas de dimensión finita, los campos de Jacobi verifican una serie de propiedades relevantes que pasamos a comentar.

Sea n la dimensión de Δ . Observemos, en primer lugar, que la ecuación de Jacobi es lineal en Y . Por lo tanto, el conjunto de campos de Jacobi a lo largo de c es un espacio vectorial sobre \mathbb{R} de dimensión $2n$, que notaremos \mathcal{F}_J .

Por otra parte, si $Y(s)$ es un campo de Jacobi a lo largo de $c(s)$, el campo $Z(s) = Y(rs)$ también es un campo de Jacobi pero a lo largo de la geodésica $c(rs)$. Este campo verifica que $Z(0) = Y(0)$ y $\frac{\nabla}{ds} Z(s_0) = r \frac{\nabla}{ds} Y(0)$.

Además, debido a la naturaleza de la ecuación de Jacobi, un campo de Jacobi Y sobre una geodésica queda determinado por su valor y el de su derivada covariante $\frac{\nabla}{\partial s}Y$ en un punto arbitrario de la geodésica.

Si X e Y son campos de Jacobi, se verifica que:

$$\begin{aligned} \frac{d}{ds} \left\{ \left\langle \frac{\nabla}{\partial s}X, Y \right\rangle - \left\langle \frac{\nabla}{\partial s}Y, X \right\rangle \right\} &= \left\langle \frac{\nabla^2}{\partial s^2}X, Y \right\rangle - \left\langle \frac{\nabla^2}{\partial s^2}Y, X \right\rangle = \\ &= -\langle R(c', X)c', Y \rangle + \langle R(c', Y)c', X \rangle = 0, \end{aligned}$$

De esta manera, podemos afirmar que:

$$\left\langle \frac{\nabla}{\partial s}X, Y \right\rangle - \left\langle \frac{\nabla}{\partial s}Y, X \right\rangle = \text{constante}$$

Si tomamos $s = c$, $X = Y = 0$, resulta:

$$\left\langle \frac{\nabla}{\partial s}X, Y \right\rangle - \left\langle \frac{\nabla}{\partial s}Y, X \right\rangle = 0. \quad (\text{A.4})$$

Proposición A.9.1. *Sea Y un campo de Jacobi a lo largo de la geodésica $c : [\alpha, \beta] \rightarrow M$. Entonces*

$$\langle Y, c' \rangle = a + bs$$

donde $a, b \in \mathbb{R}$ son constantes.

Demostración. De la compatibilidad con la métrica y considerando las relaciones de simetría del tensor curvatura, véase O'Neil [32], resulta:

$$\frac{d^2}{ds^2} \langle Y, c' \rangle = \left\langle \frac{\nabla^2 Y}{ds^2}, c' \right\rangle = -\langle R(c', Y)c', c' \rangle = 0$$

Por lo tanto, $\langle Y, c' \rangle$ es una función lineal de s .

□

La proposición anterior informa que el conjunto de todos los campos de Jacobi normales, esto es ortogonales a c' , que obtenemos fijando $a = b = 0$, forman un subespacio de $\Gamma_{\mathcal{J}}$ de dimensión $2n-2$, y los campos de Jacobi tangenciales lo forman de dimensión 2. Así podemos descomponer de forma única todo campo de Jacobi $Y(s)$ a lo largo de una geodésica $c(s)$ en la suma de un campo de Jacobi tangencial más la suma de un campo de Jacobi normal, esto es $Y = Y^{tan} + Y^{nor}$.

Finalmente comentar que todos los resultados sobre campos de Jacobi pueden ser reescritos en términos de geodésicas unitarias, esto es de geodésicas c tales

que $\|c'\| = 1$ ya que, tal y como hemos comentado anteriormente, si $Y(s)$ es un campo de Jacobi a lo largo de $c(s)$, entonces $J(s) = Y(rs)$ es un campo de Jacobi a lo largo de la geodésica $c(rs)$ cumpliendo $J(0) = Y(0)$ y $(\nabla/\partial s)J(0) = r (\nabla/\partial s)Y(0)$.

Sea ahora $\kappa : \mathbb{R} \rightarrow \mathbb{R}$ una función continua real y consideremos la ecuación diferencial ordinaria:

$$f'' + \kappa f = 0 \quad (\text{A.5})$$

Denotemos por \mathbf{S}_κ la solución de esta ecuación que verifica $\mathbf{S}_\kappa(0) = 0$ y $\mathbf{S}'_\kappa(0) = 1$. Sea \mathbf{C}_κ la solución cumpliendo las condiciones iniciales $\mathbf{C}_\kappa(0) = 1$ y $\mathbf{C}'_\kappa(0) = 0$. En el caso en que κ sea una constante real que notaremos K , solucionando la ecuación diferencial A.5 obtenemos:

$$\mathbf{S}_K(t) = \begin{cases} \frac{\sin(\sqrt{K}t)}{\sqrt{K}} & \text{si } K > 0, \\ t & \text{si } K = 0, \\ \frac{\sinh(\sqrt{-K}t)}{\sqrt{-K}} & \text{si } K < 0, \end{cases} \quad (\text{A.6})$$

y $\mathbf{C}_K = \mathbf{S}'_K$.

Dado un campo de Jacobi normal, esto es ortogonal a c' , Y , introduzcamos:

$$f_\kappa = \|Y\|(0) \mathbf{C}_\kappa + \|Y'\|(0) \mathbf{S}_\kappa \quad (\text{A.7})$$

y recordemos la definición de curvatura Riemanniana seccional:

Definición A.9.3. Sea $p \in \Delta$ y X e $Y \in T_p(\Delta)$ linealmente independientes. Introducimos:

$$K(X, Y) = \frac{\langle R(X, Y)X, Y \rangle}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}.$$

donde $R(X, Y)$ es el tensor de curvatura de la conexión de Levi-Civita, ∇ . Diremos que $K(X, Y)$ es la *curvatura Riemanniana seccional* de la 2-sección determinada por X e Y .

Con las notaciones anteriores, sean $a(s) \leq K \leq A(s)$ cotas inferior y superior de la curvatura Riemanniana seccional a lo largo de la geodésica c_p . Podemos enunciar ahora los siguientes teoremas de comparación.

Teorema A.9.2 (Primer teorema de comparación de Rauch). *Sea Y un campo de Jacobi normal a lo largo de una geodésica unitaria $c(s)$.*

Entonces, siempre que Y no se anule y $f_A(t) > 0$, $\forall 0 < t < s$, se verifican las siguientes desigualdades:

$$\left(\frac{\|Y\|(s)}{f_A(s)} \right)' \geq 0, \quad \frac{\|Y\|'(s)}{\|Y\|(s)} \geq \frac{f_A'(s)}{f_A(s)}, \quad \|Y\|(s) \geq f_A(s).$$

donde la segunda desigualdad puede ser reescrita según:

$$f_A(s) \langle \frac{\nabla}{\partial s} Y(s), Y(s) \rangle \geq f_A'(s) \langle Y(s), Y(s) \rangle.$$

Demostración. Sea $0 < t < s$. Entonces:

$$\begin{aligned} \|Y\|''(s) &= \left\{ \frac{1}{\|Y\|} \langle \frac{\nabla}{\partial s} Y, Y \rangle \right\}'(s) = \frac{1}{\|Y\|} \langle -R(c', Y)c', Y \rangle(s) \\ &\quad + \frac{1}{\|Y\|^3} \left\{ \langle \frac{\nabla}{\partial s} Y, \frac{\nabla}{\partial s} Y \rangle \langle Y, Y \rangle - \langle \frac{\nabla}{\partial s} Y, Y \rangle^2 \right\}(s) \geq \\ &\geq -A(s)\|Y\|(s) \end{aligned}$$

en virtud de la desigualdad de Cauchy-Schwarz. Así podemos afirmar que:

$$\|Y\|''(s) + A(s)\|Y\|(s) \geq 0,$$

Por otra parte, como f_A no se anula:

$$(f_A(s)\|Y\|'(s) - f_A'(s)\|Y\|(s))' \geq 0,$$

ya que $f_A'' + Af_A = 0$. Si recordamos que $\|Y\|(0) = f_A(0)$ y $\|Y\|'(0) = f_A'(0)$ e integramos de 0 a s se obtienen las desigualdades del enunciado.

Proposición A.9.3. Sea Y un campo de Jacobi normal a lo largo de una geodésica unitaria $c(t)$ y X un campo normal sobre c cumpliendo $X(0) = Y(0) = 0$ y $X(s) = Y(s)$. Entonces $\psi_Y(s) \leq \psi_X(s)$, donde:

$$\psi_X(s) \equiv \int_0^s \left\{ \langle \frac{\nabla}{\partial t} X, \frac{\nabla}{\partial t} X \rangle - \langle R(c', X)c', X \rangle \right\} dt.$$

Además $\psi_Y(s) = \psi_X(s)$ si y sólo si $X = Y$.

Demostración. Sean Y_1, Y_2, \dots, Y_{n-1} campos de Jacobi linealmente independientes que se anulen en $t = 0$. Podemos escribir:

$$X = \sum_{i=1}^{n-1} f_i Y_i, \quad Y = \sum_{i=1}^{n-1} a_i Y_i,$$

donde a_i son constantes. De la expresión anterior:

$$\begin{aligned}
\langle \frac{\nabla}{\partial t} X, \frac{\nabla}{\partial t} X \rangle &= \langle \sum_{i=1}^{n-1} f'_i Y_i + \sum_{i=1}^{n-1} f_i \frac{\nabla}{\partial t} Y_i, \sum_{j=1}^{n-1} f'_j Y_j + \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle \\
&= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle + 2 \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle \\
&\quad + \langle \sum_{i=1}^{n-1} f_i \frac{\nabla}{\partial t} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle, \\
\langle R(c', X) c', X \rangle &= - \sum_{i=1}^{n-1} f_i \langle R(c', Y_i) c', X \rangle,
\end{aligned}$$

y:

$$\begin{aligned}
dt \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle &= \langle \sum_{i=1}^{n-1} f'_i Y_i + \sum_{i=1}^{n-1} f_i \frac{\nabla}{\partial t} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle \\
&\quad + \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{\partial t} Y_j + \sum_{j=1}^{n-1} f_j \frac{\nabla^2}{\partial t^2} Y_j \rangle \\
&= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle + \langle \sum_{i=1}^{n-1} f_i \frac{\nabla}{\partial t} Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle \\
&\quad + \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{\partial t} Y_j \rangle - \langle \sum_{i=1}^{n-1} f_i Y_i, \sum_{j=1}^{n-1} f_j R(c', Y_j) c' \rangle.
\end{aligned}$$

Si recopilamos toda la información, se tiene que:

$$\begin{aligned}
\langle \frac{\nabla}{\partial t} X, \frac{\nabla}{\partial t} X \rangle - \langle R(c', X) c', X \rangle &= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle \\
&\quad + \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle + dt \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle - \langle X, \sum_{j=1}^{n-1} f'_j \frac{\nabla}{\partial t} Y_j \rangle \\
&= \langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \rangle + dt \langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \rangle \\
&\quad + \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} f'_i f_j \left\{ \langle Y_i, \frac{\nabla}{\partial t} Y_j \rangle - \langle Y_j, \frac{\nabla}{\partial t} Y_i \rangle \right\}.
\end{aligned}$$

Pero el último término de la expresión anterior es cero debido a (A.4). De esta manera:

$$\psi_X(s) = \int_0^s \left\langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \right\rangle dt + \left\langle X, \sum_{j=1}^{n-1} f_j \frac{\nabla}{\partial t} Y_j \right\rangle(s).$$

Igualmente:

$$\psi_Y(s) = \int_0^s \left\langle \sum_{i=1}^{n-1} a'_i Y_i, \sum_{j=1}^{n-1} a'_j Y_j \right\rangle dt + \left\langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{\partial t} Y_j \right\rangle(s) = \left\langle Y, \sum_{j=1}^{n-1} a_j \frac{\nabla}{\partial t} Y_j \right\rangle(s),$$

pues a_i son constantes. Observemos ahora que, al ser $f_i(s) = a_i$:

$$\psi_X(s) = \int_0^s \left\langle \sum_{i=1}^{n-1} f'_i Y_i, \sum_{j=1}^{n-1} f'_j Y_j \right\rangle dt + \psi_Y(s),$$

Esto implica:

$$\psi_Y(s) \leq \psi_X(s),$$

Finalmente observar que la igualdad se asume únicamente si $f'_i = 0$, para $i = 1, \dots, n-1$, esto es, $X = Y$.

□

Teorema A.9.4. *Sea Y un campo de Jacobi normal a lo largo de la geodésica unitaria $c(s)$ que se anule en $s = 0$. Supongamos además que para cualquier campo normal Z a lo largo de c se verifique:*

$$a(s) \leq \frac{\langle R(c', Z), Z \rangle}{\|Z\|^2},$$

Entonces:

$$\frac{\|Y\|'(s)}{\|Y\|(s)} \leq \frac{f'_a(s)}{f_a(s)} \quad \left(\frac{\|Y\|(s)}{f_a(s)} \right)' \leq 0, \quad \|Y\|(s) \leq f_a(s),$$

donde la segunda desigualdad puede ser reescrita como:

$$f_a(s) \left\langle \frac{\nabla}{\partial s} Y(s), Y(s) \right\rangle \leq f'_a(s) \langle Y(s), Y(s) \rangle$$

Demostración. Introducimos:

$$u(s) = \langle Y, Y \rangle(s), \quad v(s) = f_a^2(s)$$

y:

$$\mu(s) = \frac{\psi_Y(s)}{u(s)}, \quad \nu(s) = \frac{\int_0^s ((f'_a)^2 - af_a^2) dt}{v(s)}$$

Observemos que:

$$\int_0^s ((f'_a)^2 - af_a^2) dt = (f'_a f_a)(s) - (f'_a f_a)(0) = (f'_a f_a)(s)$$

pues $f_a f_a'' + af_a^2 = 0$ y $f_a(0) = 0$. Entonces:

$$\frac{du}{ds} = 2\langle \frac{\nabla}{\partial s} Y, Y \rangle = 2\psi_Y = 2\mu u, \quad \frac{dv}{ds} = 2f'_a f_a = 2\nu v$$

Resolviendo las anteriores ecuaciones diferenciales, se tiene que:

$$u(s) = u(\epsilon) \exp\{2 \int_\epsilon^s \mu(t) dt\}, \quad v(s) = v(\epsilon) \exp\{2 \int_\epsilon^s \nu(t) dt\}$$

Si ahora utilizamos dos veces la regla de l'Hôpital:

$$\lim_{\epsilon \rightarrow 0} \frac{u(\epsilon)}{v(\epsilon)} = \frac{\|Y'\|^2(0)}{(f'_a(0))^2} = 1,$$

con lo que:

$$\frac{u(s)}{v(s)} = \exp\left\{\int_0^s (\mu(t) - \nu(t)) dt\right\}$$

Veamos que $\mu(t) \leq \nu(t)$. Sea W un campo vectorial paralelo a lo largo de $c(t)$ tal que $W(s) = Y(s)$. Introduzcamos un nuevo campo Z dado por $Z = f_a \frac{W}{f_a(s)}$. Observemos que $Z(0) = 0$ y $Z(s) = W(s) = Y(s)$. De esta manera, la proposición anterior informa que:

$$\psi_Y(s) \leq \psi_Z(s)$$

Si ahora tenemos en cuenta que:

$$\begin{aligned} \psi_Z(s) &= \int_0^s \left\{ \left\langle f'_a \frac{W}{f_a(s)}, f'_a \frac{W}{f_a(s)} \right\rangle - \left\langle R(c', f_a \frac{W}{f_a(s)}) c', f_a \frac{W}{f_a(s)} \right\rangle \right\} dt \\ &\leq \frac{\langle W, W \rangle}{(f_a(s))^2} \int_0^s \{ (f'_a)^2 - a f_a^2 \} dt = u(s) \nu(s) \end{aligned}$$

obtenemos que:

$$\mu(s) = \frac{\psi_Y(s)}{u(s)} \leq \nu(s)$$

Así, $u(s)/v(s)$ es monótona decreciente con $u(s)/v(s) \leq 1$ de donde se obtienen las desigualdades del enunciado. \square

Proposición A.9.5. *Sea nuevamente Y un campo de Jacobi normal a lo largo de la geodésica unitaria $c(t)$ de manera que $Y(0) = 0$. Supongamos que a y A son cotas inferior y superior de la curvatura a lo largo de $c(t)$. Sea k una función continua. Entonces, si $\mathbf{S}_A > 0$ y $\mathbf{S}'_\kappa > 0$ se verifica que:*

$$\|Y'\|(t) \leq \begin{cases} \|Y\|(t) \cdot \left(\frac{2 \cdot \mathbf{S}'_\kappa}{\mathbf{S}_\kappa} - \frac{\mathbf{S}'_A}{\mathbf{S}_A} \right) (t) & \text{si } \kappa \leq \frac{1}{2}(a + A) \\ \|Y'(0)\| \cdot \mathbf{S}'_\kappa(t) & \text{si } \kappa \geq \frac{1}{2}(a + A) \end{cases} \quad (\text{A.8})$$

Demostración. Sea f_κ definida según (A.7) y Z un campo vectorial paralelo a lo largo de la geodésica $c(t)$ tal que $\|Z\| = 1$. Entonces:

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle(0) = 0$$

Por otra parte:

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' = f_\kappa \langle \kappa Y - R(c', Y) c', Z \rangle$$

Como $\langle (\kappa - R(c', \cdot) c') Y, Z \rangle$ es una forma bilineal simétrica tal que:

$$\langle (\kappa - R(c', \cdot) c') Y, Y \rangle \leq \max(A - \kappa, \kappa - a) \|Y\|^2$$

se verifica que:

$$\langle f_\kappa Y' - f'_\kappa Y, Z \rangle' \leq f_\kappa \max(A - \kappa, \kappa - a) \|Y\|$$

De aquí y recordando los dos teoremas anteriores, se tiene:

$$\begin{aligned}
& \langle f_\kappa Y' - f'_\kappa Y, Z \rangle'(t) \leq \\
& \leq \begin{cases} \frac{\|Y\|}{f_A}(s)(A - \kappa)f_\kappa f_A(t) & \text{si } \kappa \leq \frac{1}{2}(a + A) \text{ y } f_A(t) > 0, 0 < t \leq s \\ (\kappa - a)f_\kappa f_a(t) & \text{si } \kappa \geq \frac{1}{2}(a + A) \end{cases} \\
& = \begin{cases} \frac{\|Y\|}{f_A}(s)(f'_\kappa f_A - f_\kappa f'_A)'(t) & \text{si } \kappa \leq \frac{1}{2}(a + A) \text{ y } f_A(t) > 0, 0 < t \leq s \\ (f_\kappa f'_a - f'_\kappa f_a)'(t) & \text{si } \kappa \geq \frac{1}{2}(a + A) \end{cases}
\end{aligned}$$

Escojamos Z como el campo vectorial dado por $Z(s) = Y'(s)/\|Y'\|(s)$ y recordemos que por hipótesis $f'_\kappa > 0$. Si ahora integramos entre 0 y s , tenemos en cuenta la desigualdad de Cauchy–Schwarz y que $\|Y\|(s) \leq f_a(s)$ se obtiene que:

$$\|Y'\|(s) \leq \begin{cases} \|Y\| \left(2\frac{f'_\kappa}{f_\kappa} - \frac{f'_A}{f_A} \right)(s) & \kappa \leq \frac{1}{2}(a + A) \\ f'_a(s) & \kappa \geq \frac{1}{2}(a + A) \end{cases}$$

Finalmente, si recuperamos las expresiones de f_κ y de f_A la proposición queda probada. □

A.10 Campos de Killing

Finalizaremos este compendio de definiciones y resultados sobre Geometría Diferencial con los campos de Killing. Estos son de gran importancia en el capítulo 7.

En primer lugar necesitamos recordar el concepto de Derivada de Lie.

Sea Δ una variedad \mathcal{C}^∞ y p un punto de Δ . Como $T_p(\Delta)$ tiene estructura de espacio vectorial, podemos recurrir al Álgebra Lineal y definir lo que entenderemos por tensores.

Sea $T_p(\Delta)^*$ el dual topológico de $T_p(\Delta)$.

Definición A.10.1. Un *tensor r -covariante y s -contravariante* ($r, s > 0$) en $p \in \Delta$ es una aplicación multilineal:

$$(T_p(\Delta))^r \times (T_p(\Delta)^*)^s \longrightarrow \mathbb{R}$$

donde $(T_p(\Delta))^r$ indica r copias de $T_p(\Delta)$ y $(T_p(\Delta)^*)^s$ denota s copias de $T_p(\Delta)^*$.

Es inmediato verificar que el conjunto de tensores r -covariantes y s -contravariantes en un punto tiene estructura de espacio vectorial. Denotaremos este espacio vectorial por $\mathcal{T}^{s,r}(T_p(\Delta))$.

Casos especiales son aquellos en los que $s = 0$, $r = 0$ y $s = r = 0$. En los dos primeros dispondremos de el espacio vectorial de los r -cotensores, $\mathcal{T}^{0,r}(T_p(\Delta))$, y de los s -contratensores, $\mathcal{T}^{s,0}(T_p(\Delta))$. Para el tercero, definiremos $\mathcal{T}^{0,0}(T_p(\Delta))$ como el cuerpo K sobre el que \mathbf{E} es espacio de Banach.

Finalmente observemos que un vector en p es un 1-contratensor en p . Los 1-cotensores en p se denominan 1-formas en p .

Podemos definir ahora los campos tensoriales de clase \mathcal{C}^∞ de igual manera a como a partir de los vectores tangentes definimos los campos tangentes.

Definición A.10.2. Sea $A \subset \Delta$. Diremos que una aplicación α :

$$A \ni p \longmapsto \alpha_p \in \mathcal{T}^{s,r}(T_p(\Delta))$$

es un s, r *campo tensorial de dominio A* .

Definición A.10.3. Si en la definición anterior A es abierto y para todos los conjuntos de s campos vectoriales \mathcal{C}^∞ en A , Y_1, \dots, Y_r , y de s 1-formas en A , w_1, \dots, w_s , la función:

$$A \ni p \longmapsto [\alpha(Y_1, \dots, Y_r; w_1, \dots, w_s)](p) = \alpha_p(w_1(p), \dots, w_s(p); Y_1(p), \dots, Y_r(p))$$

es de clase \mathcal{C}^∞ diremos que α es un *campo tensorial \mathcal{C}^∞ de dominio A* . Denotaremos este conjunto como $\mathcal{T}^{s,r}(A)$, y es inmediato verificar que al igual que ocurría con los tensores, $\mathcal{T}^{s,r}(A)$ tiene estructura de espacio vectorial.

Disponemos ahora de los elementos necesarios para introducir la derivada de Lie.

Sean $r, s \geq 0$ y X un campo \mathcal{C}^∞ de dominio un conjunto abierto A .

Definición A.10.4. La *Derivada de Lie via X* , que denotaremos L_X , es un operador que aplica $\mathcal{T}^{s,r}(A)$ en él mismo según la siguiente definición:

- a. Si f es una función $\mathcal{C}^\infty(A)$, entonces $L_X f = D_X f$
- b. Si $Y \in \mathcal{T}^{1,0}(A)$, definimos $L_X Y = [X, Y]$, donde recordemos que el paréntesis de Lie $[X, Y] = D_X Y - D_Y X$.
- c. Si $w \in \mathcal{T}^{0,1}(A)$, $(L_X w)(Y) = Xw(Y) - w([X, Y])$, $Y \in \mathcal{T}^{1,0}(A)$.
- d. Si $\alpha \in \mathcal{T}^{s,r}(A)$, siendo $r, s > 0$, entonces $L_X \alpha$ se obtiene como solución de la ecuación:

$$L_X[\alpha(Y_1, \dots, Y_r; w_1, \dots, w_s)] = (L_X \alpha)(Y_1, \dots, Y_r; w_1, \dots, w_s) + \\ + \alpha(L_X Y_1, \dots, Y_r; w_1, \dots, w_s) + \dots + \alpha(Y_1, \dots, Y_r; w_1, \dots, L_X w_s)$$

Los campos de Killing van a ser una clase especial de campos vectoriales relacionados con la derivada de Lie. Supongamos ahora que Δ sea una variedad Riemanniana y sea g el tensor métrico.

Definición A.10.5. Sea X un campo vectorial \mathcal{C}^∞ de dominio un conjunto abierto A . Diremos que X es un campo de Killing si y sólo si $L_X g = 0$.

Los campos de Killing verifican un importante conjunto de propiedades del que únicamente destacaremos las que vamos a necesitar en el capítulo 7. Sin embargo pueden consultarse más detalles al respecto especialmente en O'Neil [32] aunque también en Petersen [33].

Proposición A.10.1. Sea (Δ, g) una variedad Riemanniana y $A \subset \Delta$ un conjunto abierto. Un campo vectorial $X \in \mathcal{C}^\infty(A)$ es de Killing si y sólo si sus curvas integrales definen isometrías.

Proposición A.10.2. El conjunto de los campos de Killing de una variedad Riemanniana (Δ, g) de dimensión finita m forma un álgebra de Lie de dimensión menor o igual que $\frac{(m+1)m}{2}$.

El primer resultado es una consecuencia directa de la definición de campo de Killing, mientras que el segundo puede encontrarse demostrado con detalle en la página 165 de Petersen [33].

Con este apartado damos por concluido este pequeño compendio de conceptos y resultados de Geometría Diferencial. Aunque ya se han ido citando a lo largo de este capítulo, más detalles sobre cualquiera de estos temas pueden consultarse en Chavel [12], Lang [26], Hicks [23], Petersen [33] o bien O'Neil [32].

Bibliografía

- [1] M. Abramovitz. *Handbook of Mathematical Functions*. Dover Publications Inc., New York, 1970.
- [2] S-I. Amari. *Differential Geometry Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1985.
- [3] S-I. Amari, O.E. Barndorff-Nielsen, R.E. Kaas, S-L. Lauritzen, and C.R. Rao. *Differential Geometry in Statistical Inference*, volume 10 of *Lecture Notes Monograph Series*. Springer-Verlag, New York, 1987.
- [4] T.M. Apostol. *Análisis Matemático*. Editorial Reverté, Barcelona, 1996.
- [5] C. Atkinson and A.F.S. Mitchel. Rao's distance measure. *Sankhyà*, 43:A: 345–365, 1981.
- [6] A.V. Balakrishnan. *Functional Analysis*, volume 10 of *Lecture Notes Monograph Series*. Springer-Verlag, New York, 1987.
- [7] A. Battacharyya. On discrimination and divergence. *Proc. 29th Indian Sci. Cong.*, Part III,13, 1943.
- [8] D. Brillinger. Necessary and sufficient conditions for a statistical problem to be invariant under a lie group. *Ann. Math. Statist.*, 34:492–500, 1963.
- [9] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematicae*, 4, 1986.
- [10] J. Burbea and J.M. Oller. The information metric for univariate linear elliptic models. *Statistics & Decisions*, 6:209–221, 1988.
- [11] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.

- [12] I. Chavel. *Riemannian Geometry. A modern introduction*. Cambridge Tracts in Mathematics. Cambridge University Press, 1993.
- [13] J.M. Corcuera. *Análisis Intrínseco de la estimación puntual*. Universitat de Barcelona. Tesis doctoral, Departament d'Estadística, 1994.
- [14] I. Csiszar. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungrica*, 2:299–318, 1967.
- [15] C.M. Cuadras and J. Fortiana. A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis*, 52,1:1–14, 1995.
- [16] A.P. Dawid. Further comments on some comments on a paper by efron. *The Annals of Statistics*, 5,6:1249, 1977.
- [17] M.L. Eaton. *Multivariate Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, 1983.
- [18] G. García and J.M. Oller. Minimum riemannian risk equivariant estimator for the univariate normal model. *Statistics and Probability Letters*, 52:109–113, 2001.
- [19] N.C. Giri. *Multivariate Statistical Inference*. Academic Press, New York, 1977.
- [20] N.C. Giri. *Group Invariance in Statistical Inference*. World Scientific, Singapore, 1996.
- [21] W.J. Hall, R.A. Wijsman, and J.K. Ghosh. The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.*, 36:575–614, 1965.
- [22] S. Helgason. *Differential Geometry, Lie Groups and Symetric Spaces*. Academic Press, 1978.
- [23] N.J. Hicks. *Notes on Differential Geometry*. Van Nostrand Reinhold, New York, 1965.
- [24] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30:509–541, 1977.
- [25] S. Kobayashi and K. Numizu. *Foundations of differential geometry*. Interscience Publishers, John Wiley, New York, 1969.

- [26] S. Lang. *Differential and Riemannian Manifolds*. Springer-Verlag, 1995.
- [27] E.L. Lehmann. A general concept of unbiasedness. *Annals Math. Statist.*, 22:587–592, 1951.
- [28] E.L. Lehmann. *Theory of point estimation. 2nd Ed.* Springer Texts in Statistics. Springer-Verlag, 1998.
- [29] P.C. Mahalanobis. On the generalized distance in statistics. *Proc. Nat. Inst. Sc. India*, 2(1):49–55, 1936.
- [30] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley, New York, 1982.
- [31] J.M. Oller and J.M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5):1562–1581, 1995.
- [32] B. O’Neil. *Semi-Riemannian Geometry, with applications to Relativity*. Pure and Applied Mathematics. Academic Press, New York, London, 1983.
- [33] P. Petersen. *Riemannian Geometry*, volume 171. Springer-Verlag, New York, 1991.
- [34] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.
- [35] C.R. Rao. Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.*, 21:24–43, 1982.
- [36] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1973.
- [37] C. Stein. Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1:197–206, 1956.
- [38] C. Stein and James W. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:361–379, 1961.
- [39] R.A. Wijsman. *Invariant Measures on Groups and Their Use in Statistics*, volume 14 of *Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, 1990.